Lecture 03

Expressiveness & Effectiveness
Basic Visualisation Techniques I

Uta Hinrichs

CS5044 – Information Visualization

University of
St Andrews

practical 1 – frequently asked questions

1. Do visualization and legend have to be in the postcard format and style as the provided visualizations?
   - You can do this as you wish. You can follow the postcard format and style, or provide a simple sketch with the legend next to it.

2. Am I required to back-up my argument with lots of additional readings on perceptual theory, etc...
   - No, a brief and concise critique according to the lecture material and book chapter is sufficient.

3. Should I comment on additional aspects in the visualization (e.g., intended themes, applied visual metaphors, or visible trends)?
   - Yes, if anything sticks out for you, you can comment on this. But you still have to stick to the word limit.

frequently asked questions

4. The visualization provided to me is already very good, and I don't know how to critique it.
   - Focus on the aspects that make it a good visualization (e.g., use of visual variables, expressiveness, effectiveness, spatial arrangement, etc...).
   - In the re-design you could try to highlight other aspects that might be of interest in the data.

5. The visualization provided to me is quite bad, and I cannot read it.
   - Describe the aspects that make it so difficult to read (e.g., use of visual variables, expressiveness, effectiveness, spatial arrangement, etc...).
   - Is there really not a single aspect that works? Is there really nothing you can decipher?
   - In the re-design you can then address these aspects that don't work.

frequently asked questions

6. Does my visualization sketch have to strictly focus on refining and improving the provided visualization, or can I come up with a sketch that focuses on other aspects (e.g., different attributes, etc…)?
   - You can provide a sketch that highlight out other aspects, but it should (1) still share at least some of the provided data attributes, and (2) you should justify your decisions carefully.
   - This type of approach makes most sense, if the provided visualization is quite good already, or if you are collecting your own data and the provided categories do not make much sense.

7. For our critique, should we comment on the design of the visualization in the context of the postcard? Some elements may not work on a small postcard but may actually work at a larger scale or with a computationally generated vis.
   - Generally, yes, the visualization should work in the context of the postcard. However, you can make a brief comment that some of the drawbacks you may encounter can be mitigated using a larger display or another presentation medium…
   - → For your own visualization design you should consider a static medium (e.g., paper).

outline for today

- Expressiveness & effectiveness principles
- Basic visualisation techniques for encoding tabular data

by the end of the lecture

- Have a solid understanding of the expressiveness and effectiveness principle and be able to critique a visualisation based on this

- Have a solid understanding of
  - Scatterplots
  - Bar charts
  - Line graphs

- and what types of attributes they can represent

expressiveness & effectiveness

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Effectiveness: Most → Least

Same

Same

Identity Channels: Categorical Attributes

Spatial region

Color hue

Motion

Shape

8

choice of marks and variables

- Your choice of marks and visual variables to encode data attributes should follow 2 principles

→ Expressiveness
→ Efficiency

expressiveness principle

- The visual encoding should express all of, and only, the information in the dataset attributes.

→ Show ordered data in a way that our perceptual system intrinsically senses as ordered

→ Show categorical data in a way that does not perceptually suggest ordering.

expressive if all we have is population size and country name?

- No!
- The choice of colours implies an ordering that is not apparent in the data
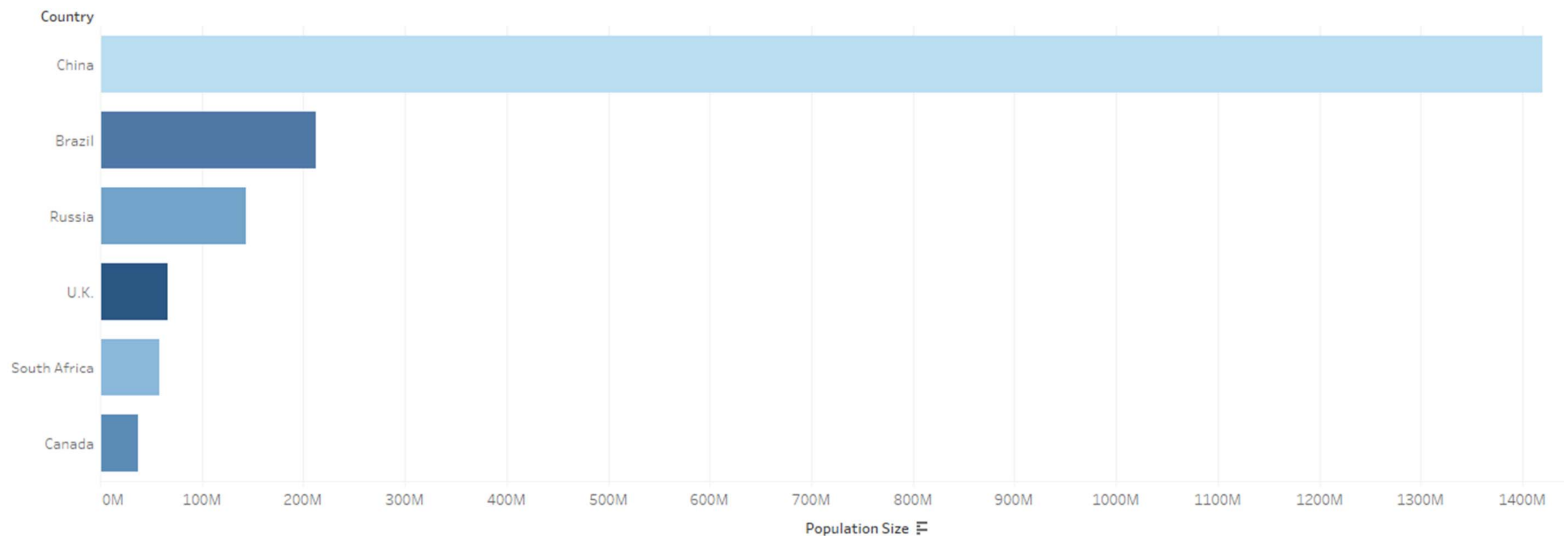
expressive if all we have is population size and country name?

- Hue is the better choice here.
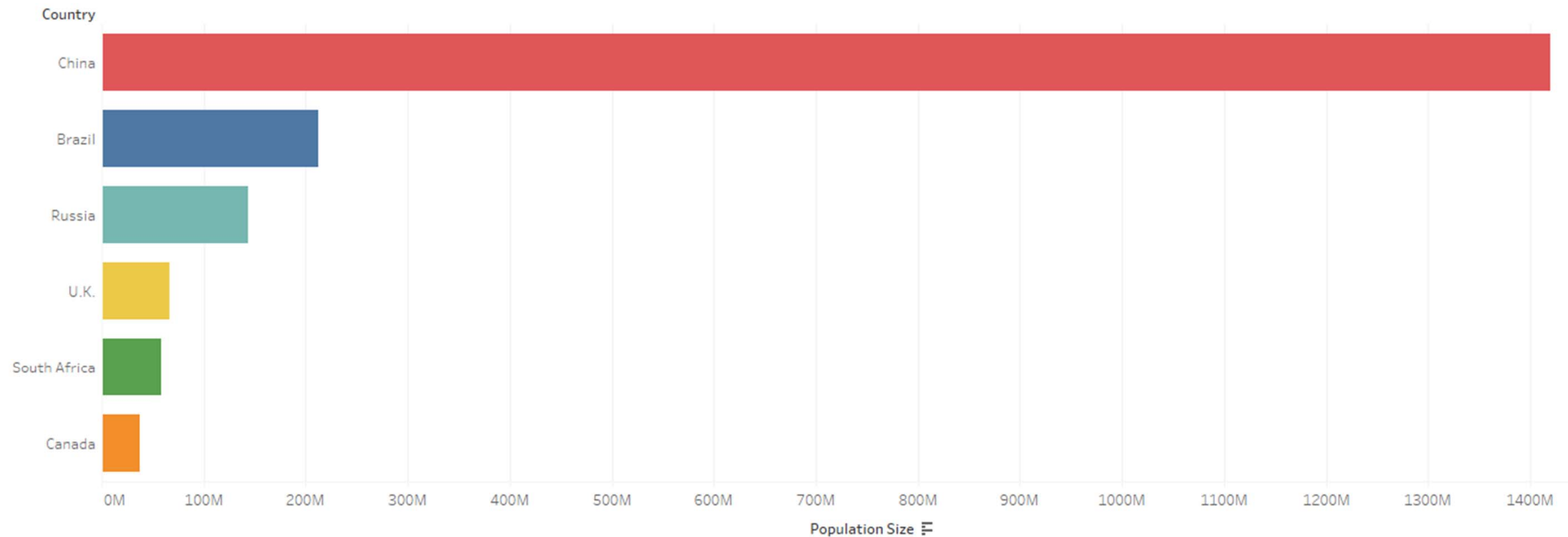
expressive: with population size, country, and area (in km²)
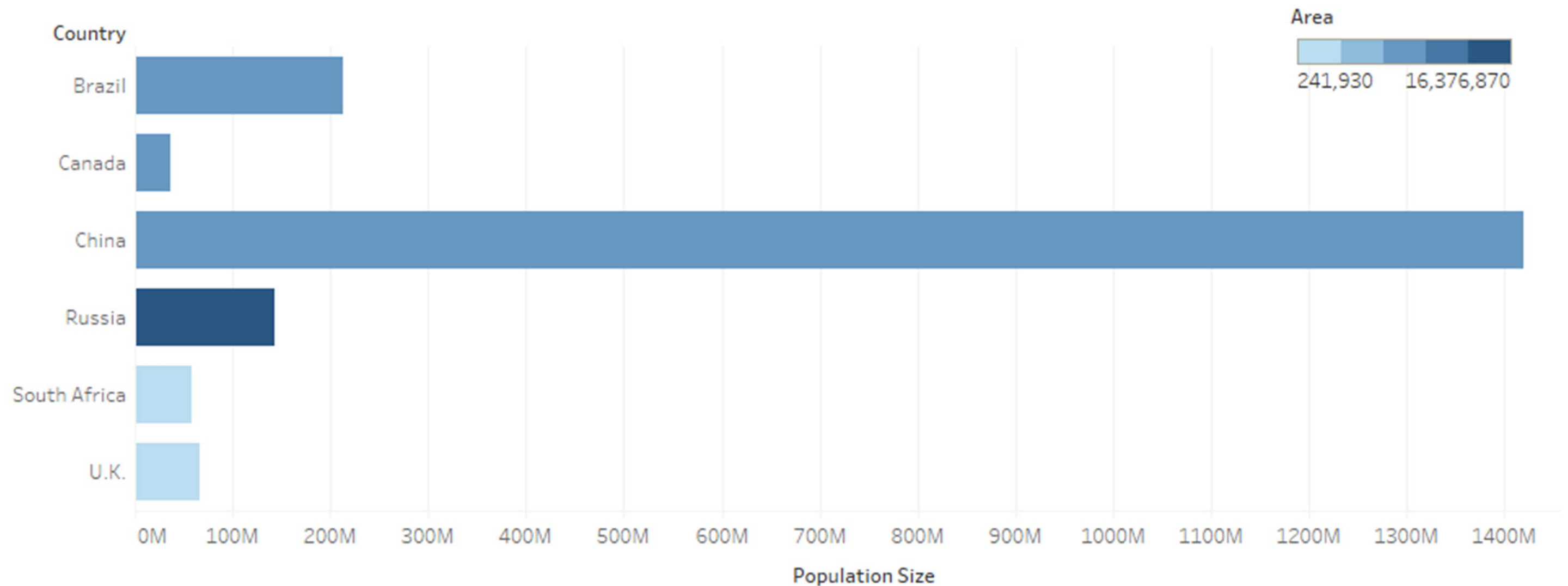
effectiveness principle

- The importance of the data attribute should be encoded with the most effective visual variable, in order to be noticeable.

→ Accuracy

→ Discriminability

→ Separability

accuracy

- How accurate is our perception of a visual variable?

- Perceived sensation (S)
- Physical intensity (I)
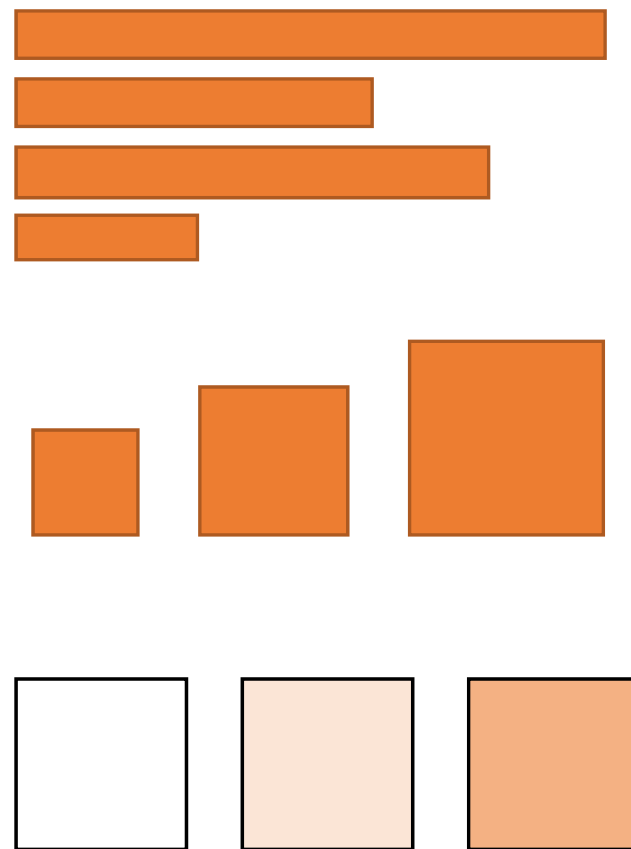
Steven's Psychophysical Power Law: $S = I^N$
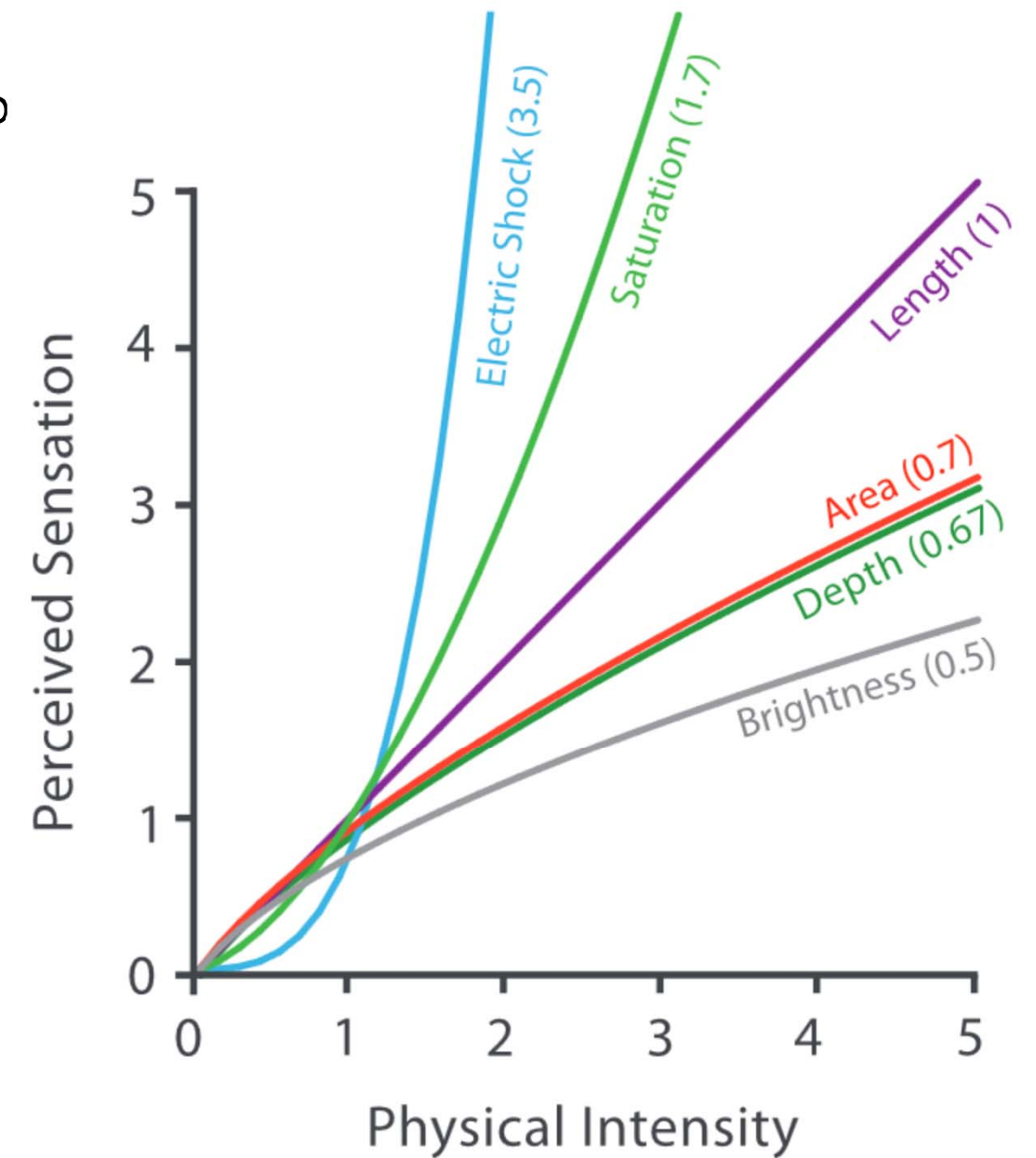
discriminability

- How many distinctions are possible?
  - Colour (hue) is highly limited (7 distinctions to be sure)
  - Position is much more versatile

- Number of distinctions can depend on context
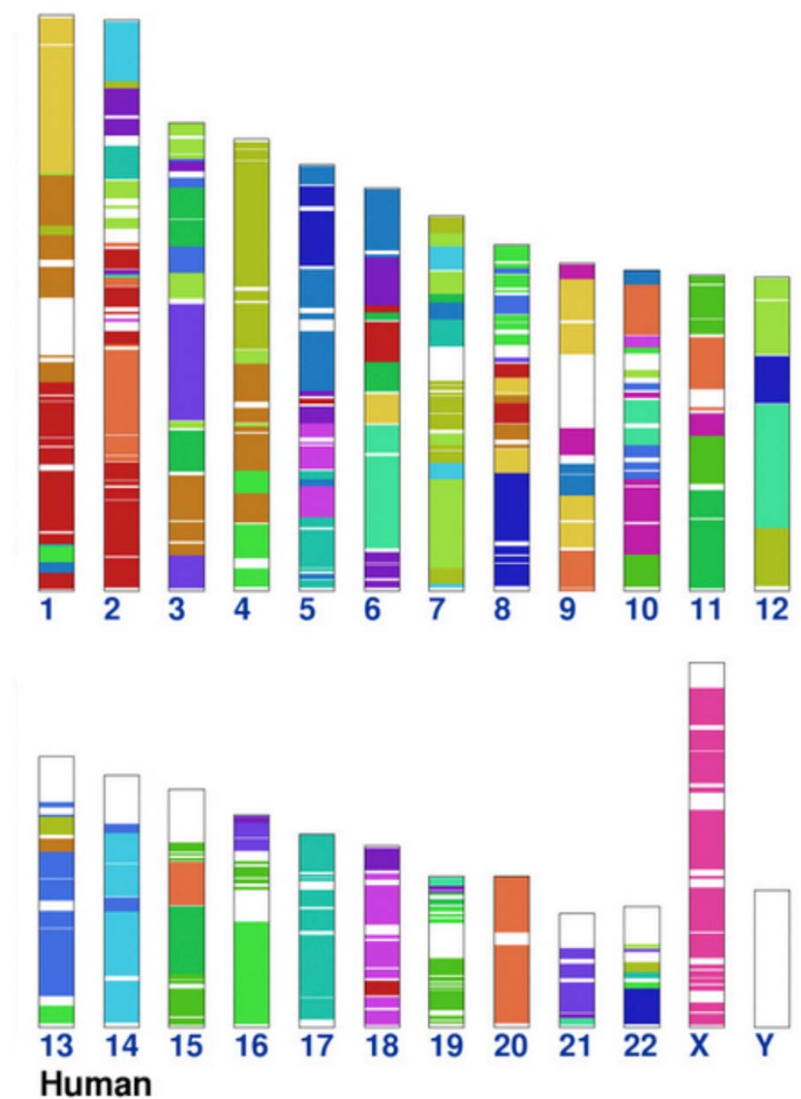
Sinha and Meller, 2007

separability

- How does the visual variable interact with other visual variables?
- Interferences between different visual variables

- Example: position + color (hue)
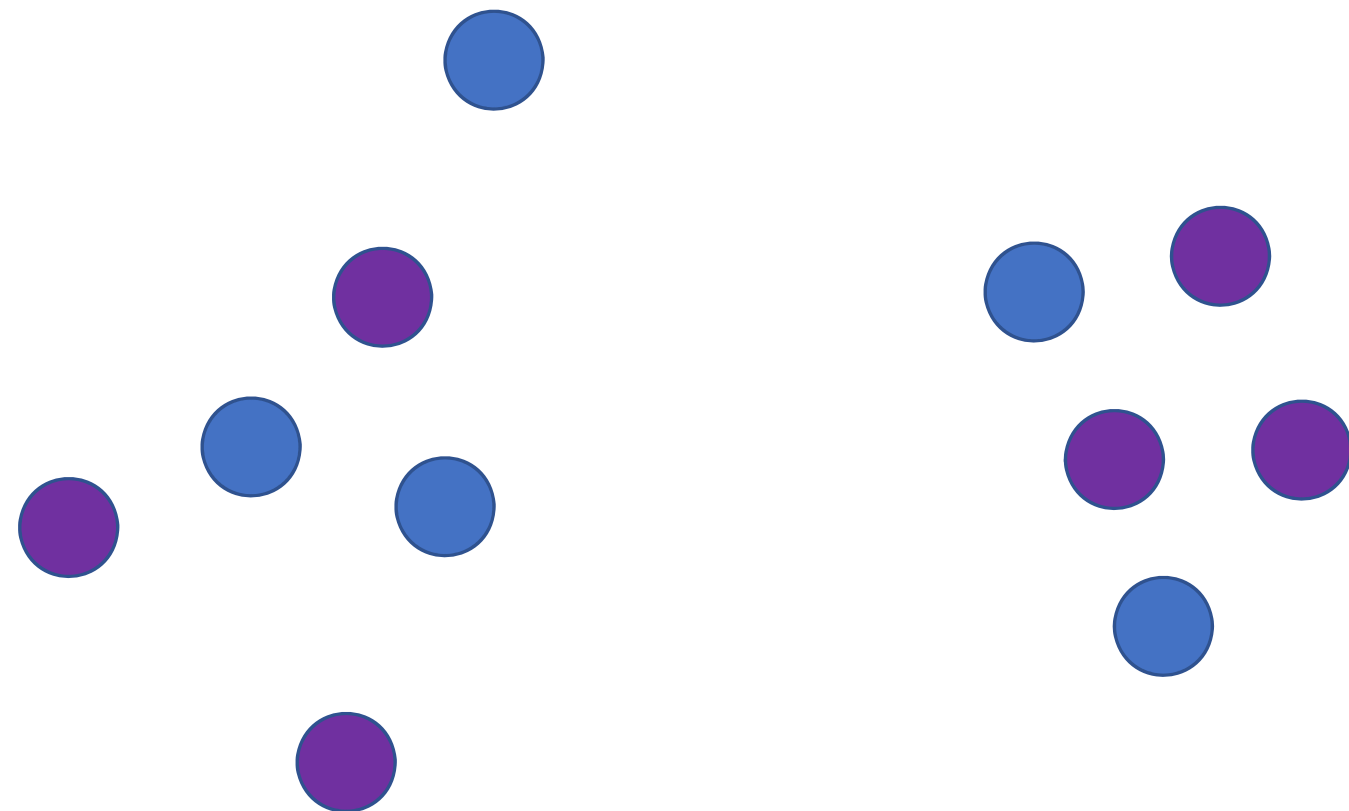
separability

- How does the visual variable interact with other visual variables?
- Interferences between different visual variables

- Example: size + color (hue)

pop out & grouping

→ Selective
  – How easy can we identify items that stand out?
  – Strong variables: position, colour, shape…
  – Often depends on the number of distractors

→ Associative
  – How can we see if items are similar; belong to the same group
  – Strongest variable here: position

things to consider

- Carefully consider expressiveness & effectiveness of your visual representation
  - If you violate these principles, you should be able to justify your decision!

- What do you want to achieve with your visualization?
  - Accurate comparisons vs. broad explorations

- Interactivity can help to ease limitations of visual variables
  - But try not to rely on it!

expressiveness & effectiveness

visualization techniques

A PERIODIC TABLE OF VISUALIZATION METHODS

classifying visualization techniques

- Based on the number of attributes
- Based on the spatial arrangement

classifying vis techniques based on # of attributes

- How many attributes/dimensions/variables does the data set have?

classifying vis techniques based on # of attributes

- Univariate data
  - Refer to just a single number (e.g., temperature, speed, time)

classifying vis techniques based on # of attributes

- Univariate data
  - Refer to just a single number or attribute (e.g., temperature, speed, time)
  - We may have many data along a single dimension

Id	Age
1	8
2	7
3	9
4	13
5	10

Total Age Distribution (n = 73)
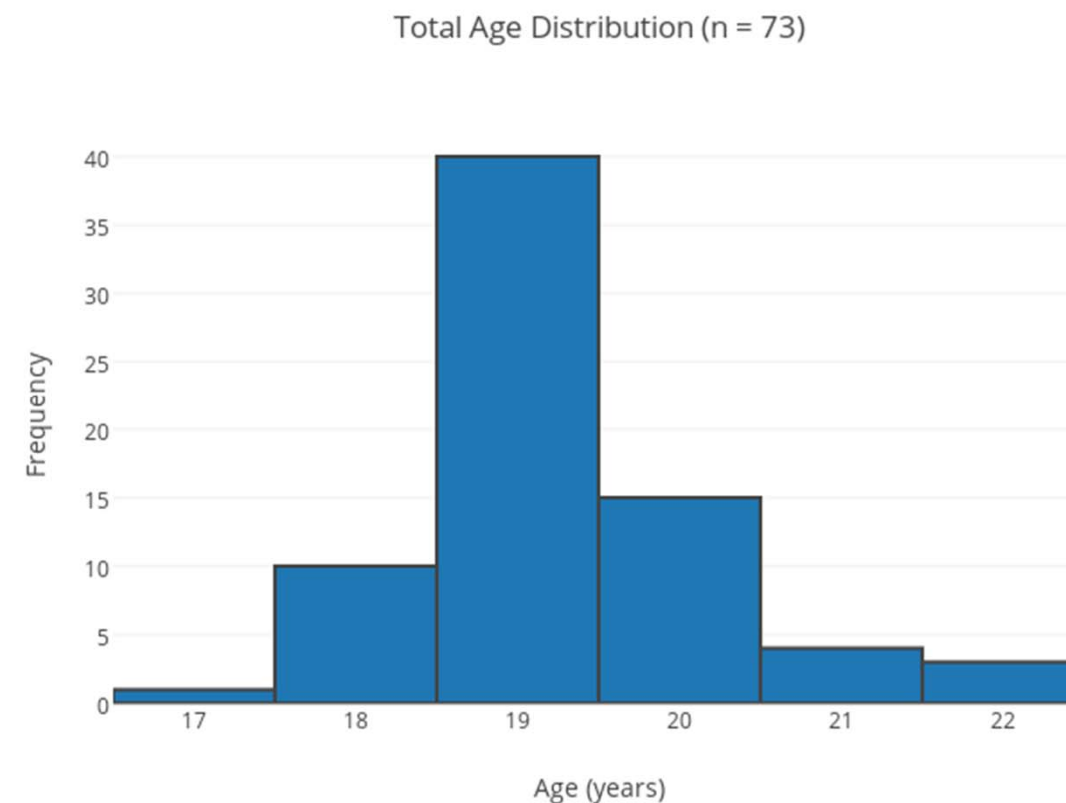
classifying vis techniques based on # of attributes

- Bivariate data
  - Data along two dimensions/variables

Id	Height	Age
1	1.10	8
2	1.00	7
3	1.20	9
4	1.70	13
5	1.20	10

classifying vis techniques based on # of attributes

- Multivariate data
  - Data along multiple dimensions

Id	Name	Height	Age	Shirt Size	Fav. Fruit
1	Apple	1.10	8	S	Banana
2	Basil	1.00	7	S	Pear
3	Clara	1.20	9	M	Durian
4	Desmond	1.70	13	L	Elderberry
5	Fanny	1.20	10	S	Lychee

classifying vis techniques based on spatial arrangement

- T. Munzner: How to arrange tabular data spatially?
  - How many dimensions do I want to show?
  - What types of attributes am I dealing with?
  - How many values do I want to show?
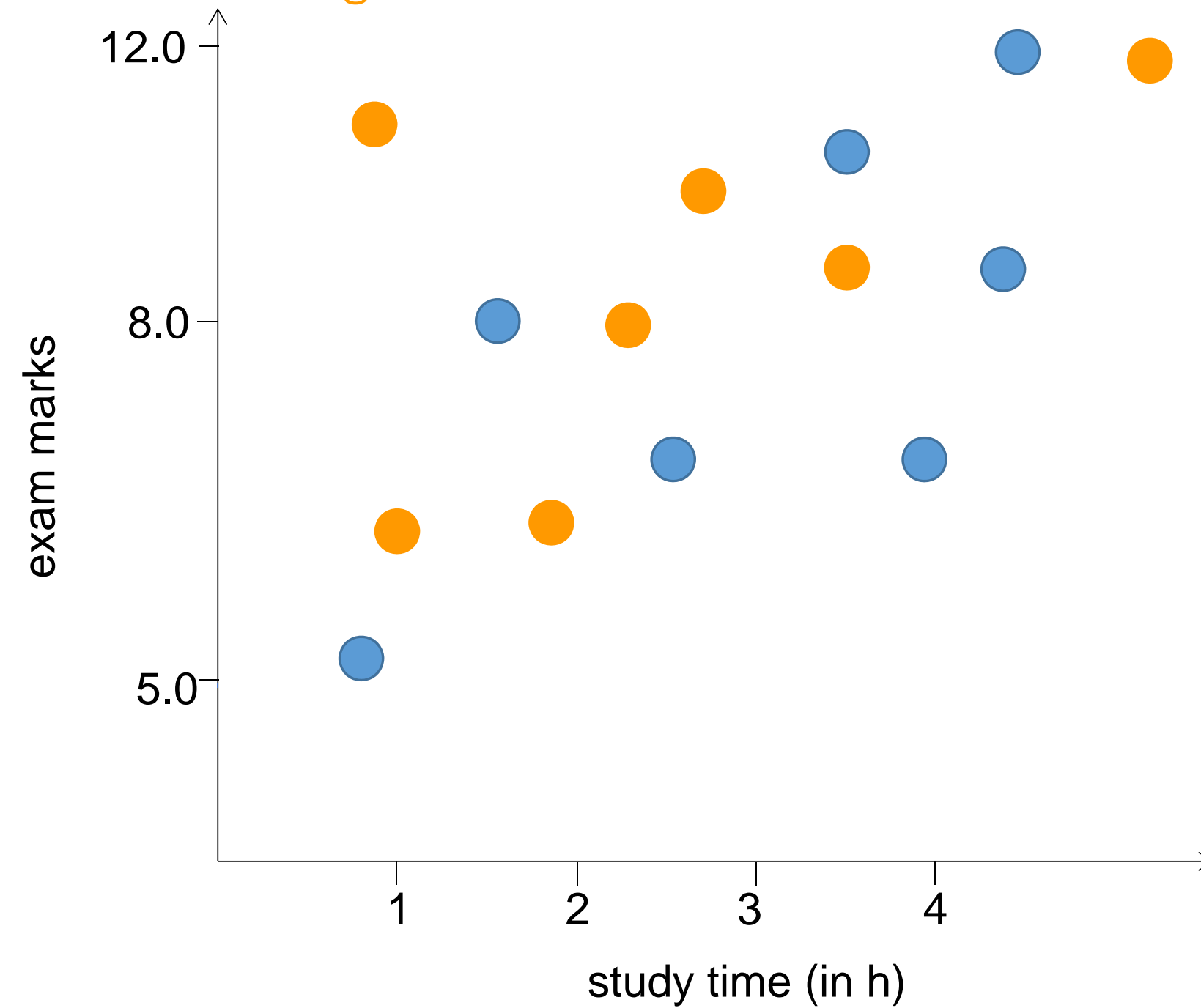
→ Choice of visualization technique

scatterplot

exam marks

study time (in h)
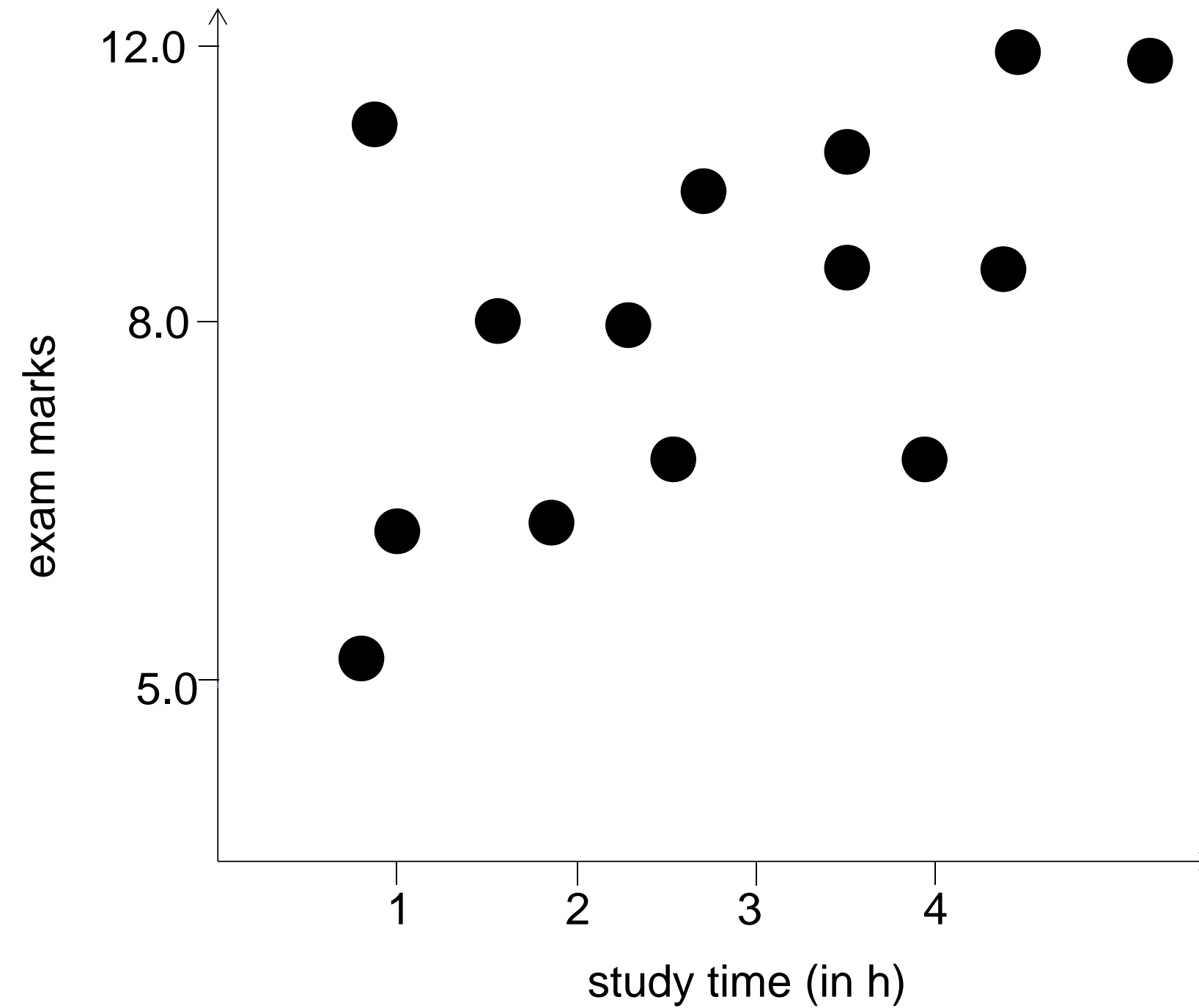
scatterplot

two quantitative attributes + one categorical attribute

scatterplot / bubble chart
three quantitative attributes + one categorical attribute

- Infant mortality

- Income per person

- Population size

- Geographic region

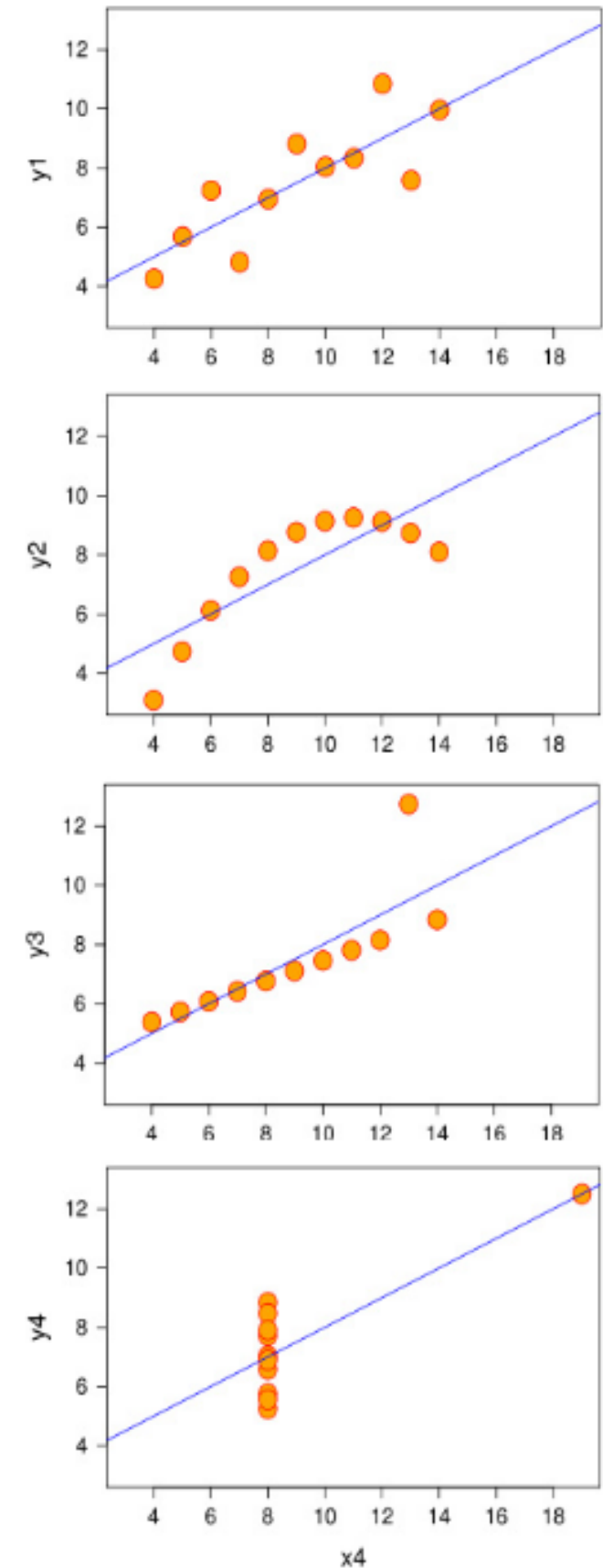Gapminder, income per person vs. infant mortality

scatterplot

Two quantitative attributes (at least)

Express quantitative values as point marks using horizontal and vertical position

Providing overviews

- Characterizing distributions
  - Outliers; extremes

- Finding correlations between two attributes

33

scatterplot

Two quantitative attributes (at least)
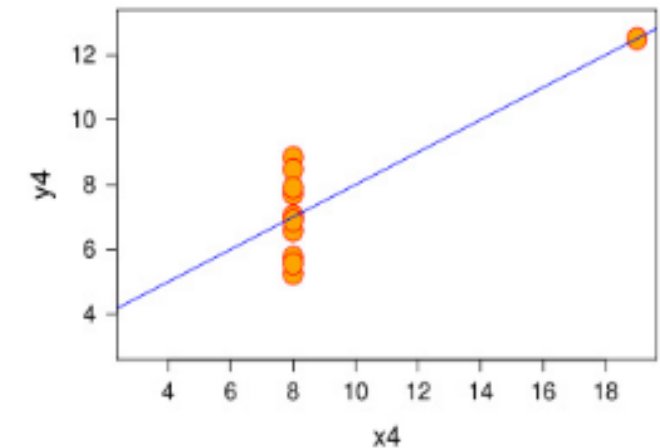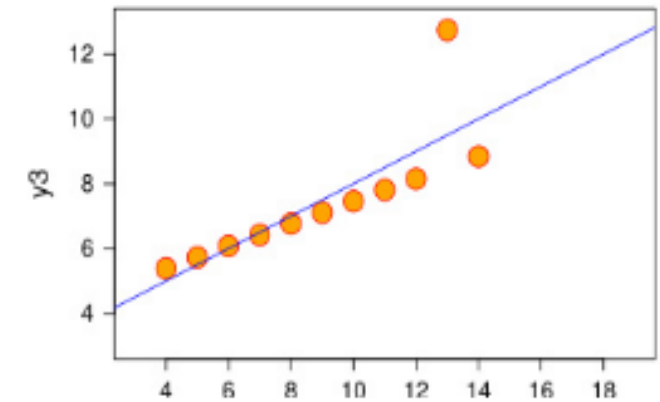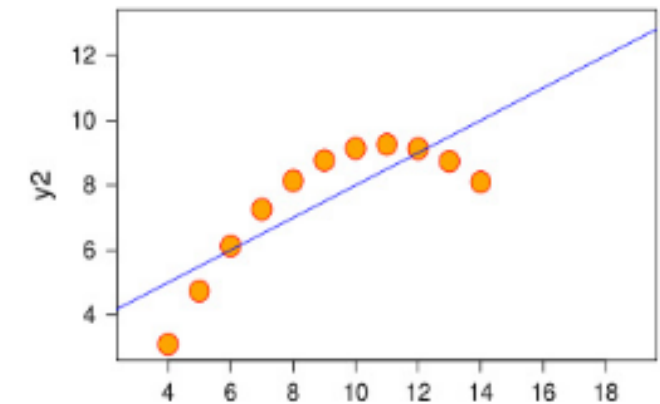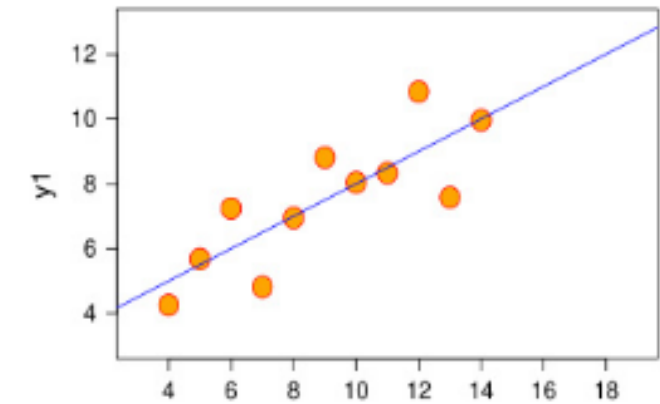
Express quantitative values as point marks using horizontal and vertical position
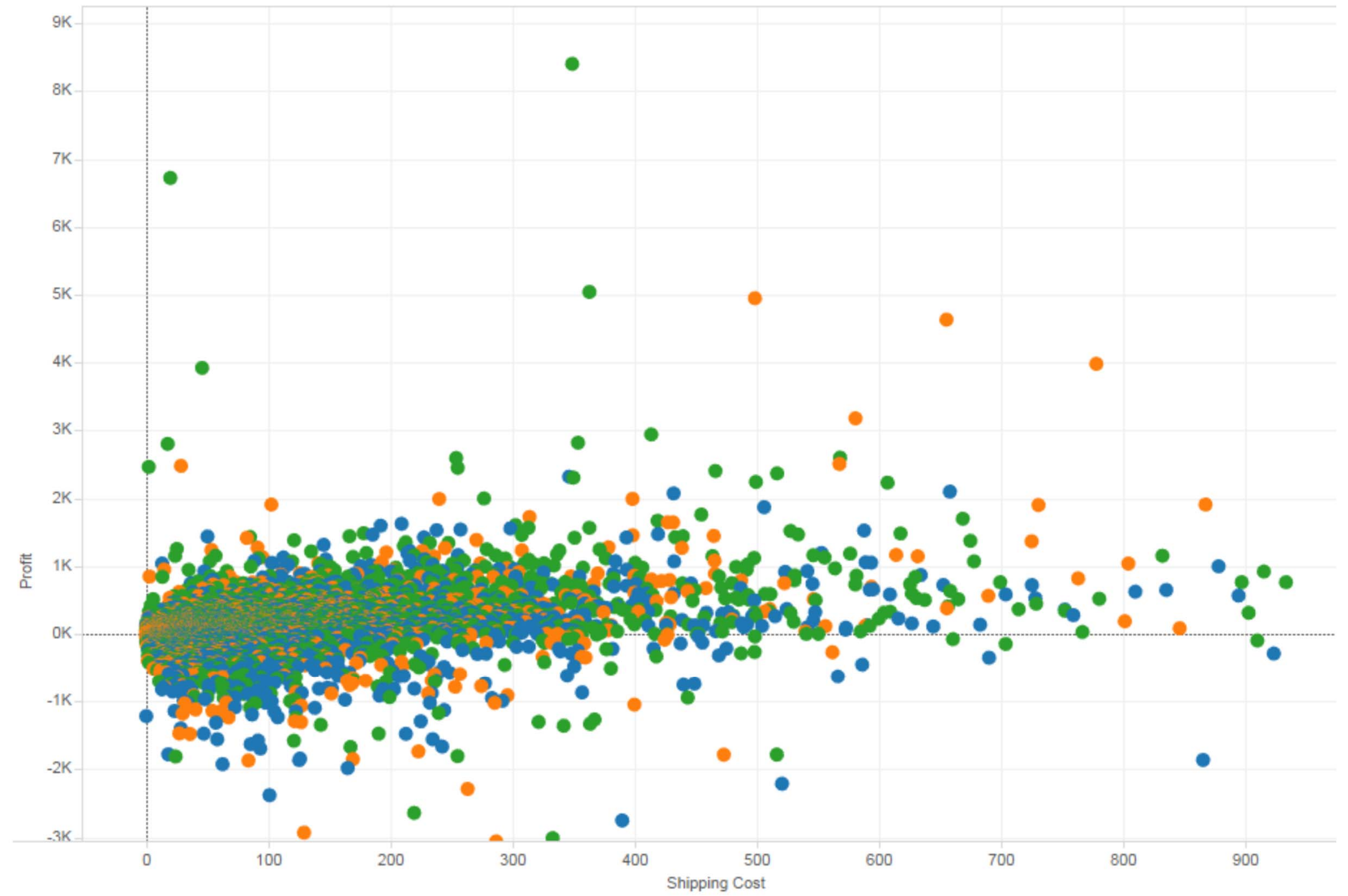
Providing overviews

- Characterizing distributions
  - Outliers; extremes
- Finding correlations between two attributes

Items in the hundreds

→ Distinguishing data points from each other becomes difficult in larger data sets
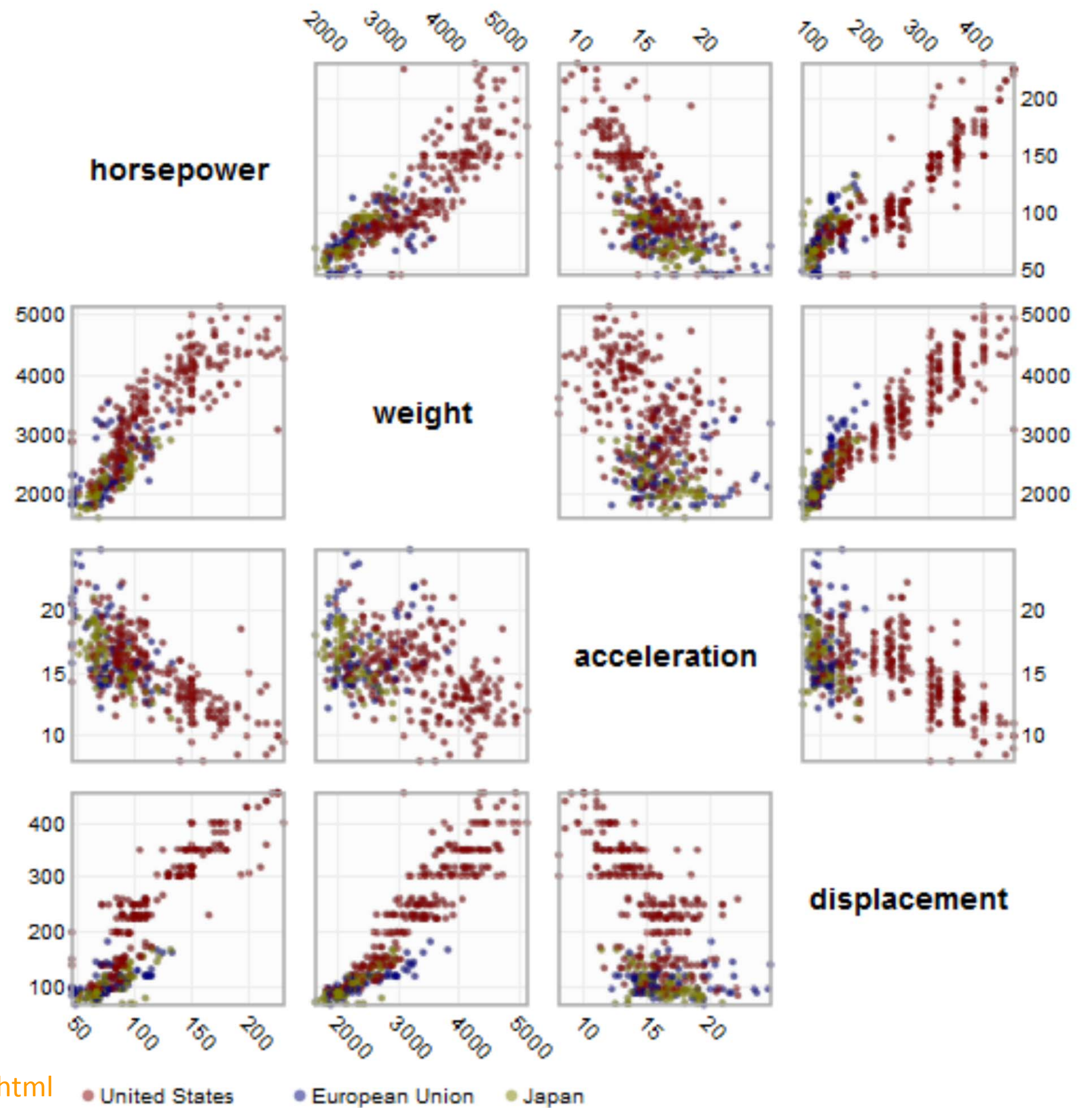
34

scatterplot

What if we want explore pairwise correlations across multiple quantitative attributes?

horsepower	weight	acceleration	displacement
50	3000	10	400
65	4000	25	300
250	3500	15	270
125	2100	17	350
200	2300	12	125

scatterplot matrix

scatterplot matrix

why ▸ Find pairwise correlations

- Trends
- Outliers

scale ▸ Attributes < 12

- Data points in the hundreds (as for scatterplots)
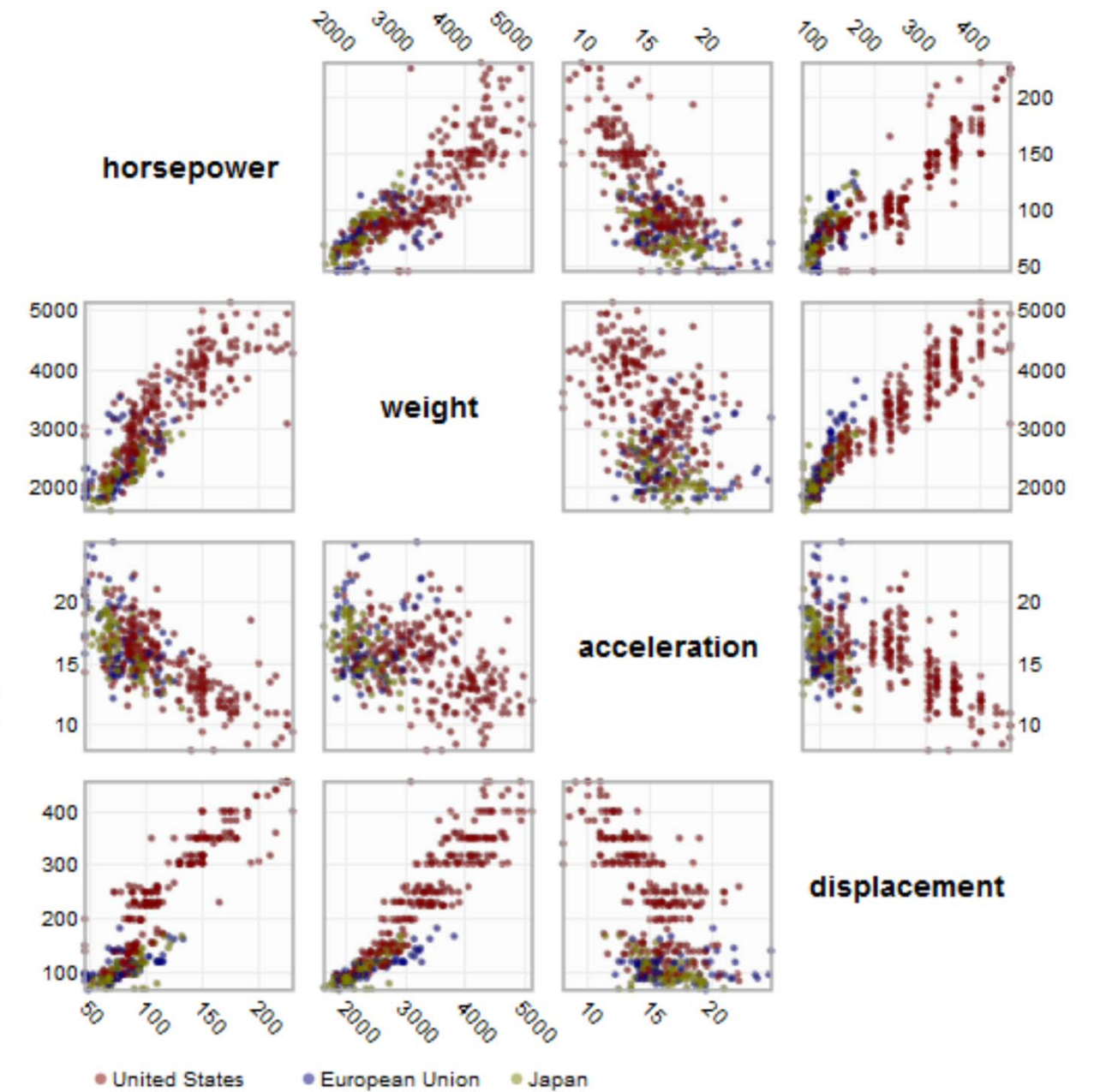
bar charts

ID	Animal type	Avg. weight
1	Capybara	78
2	Cat	13
3	Wombat	50
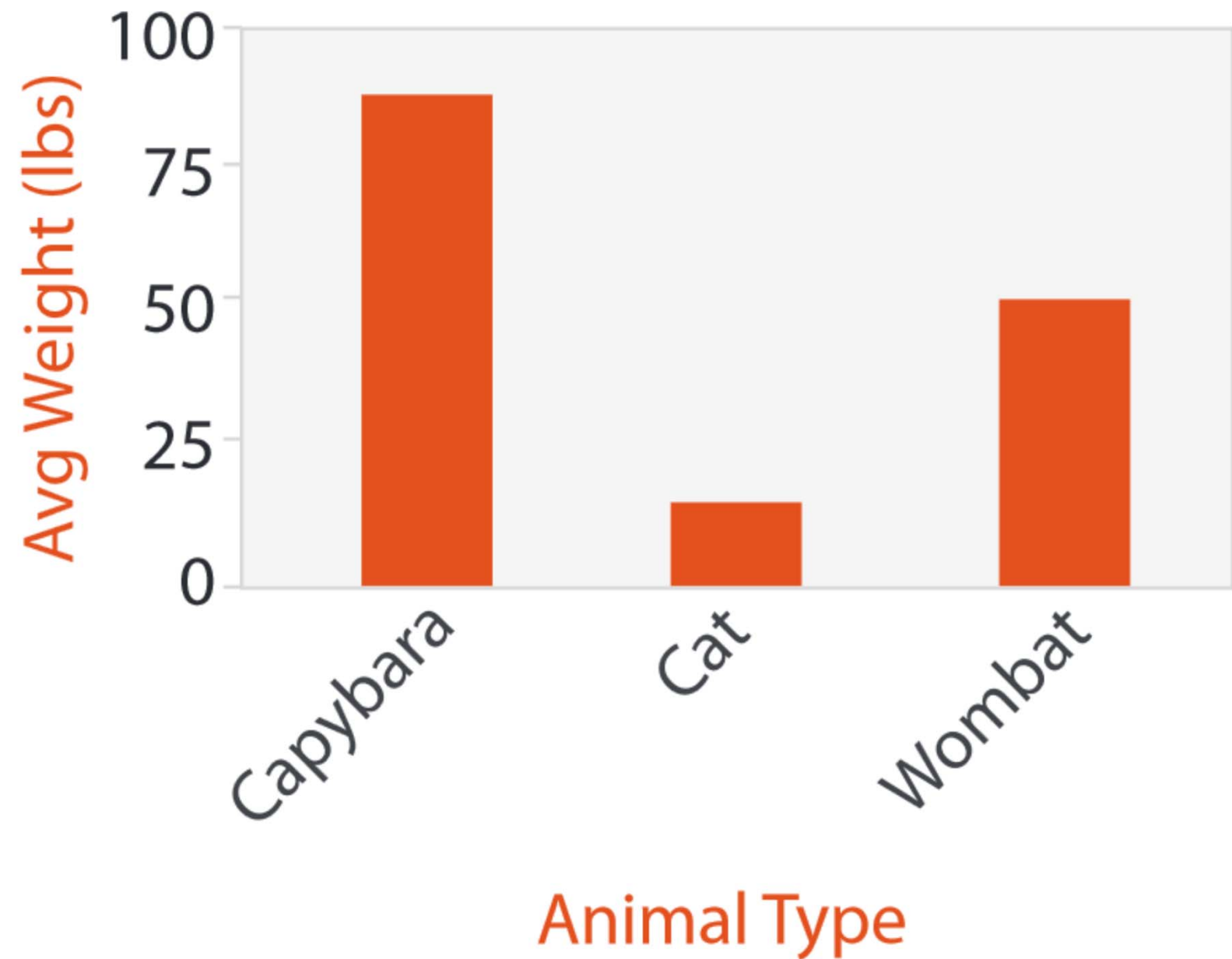
bar charts

ID	Animal type	Avg. weight
1	Capybara	78
2	Cat	13
3	Wombat	50
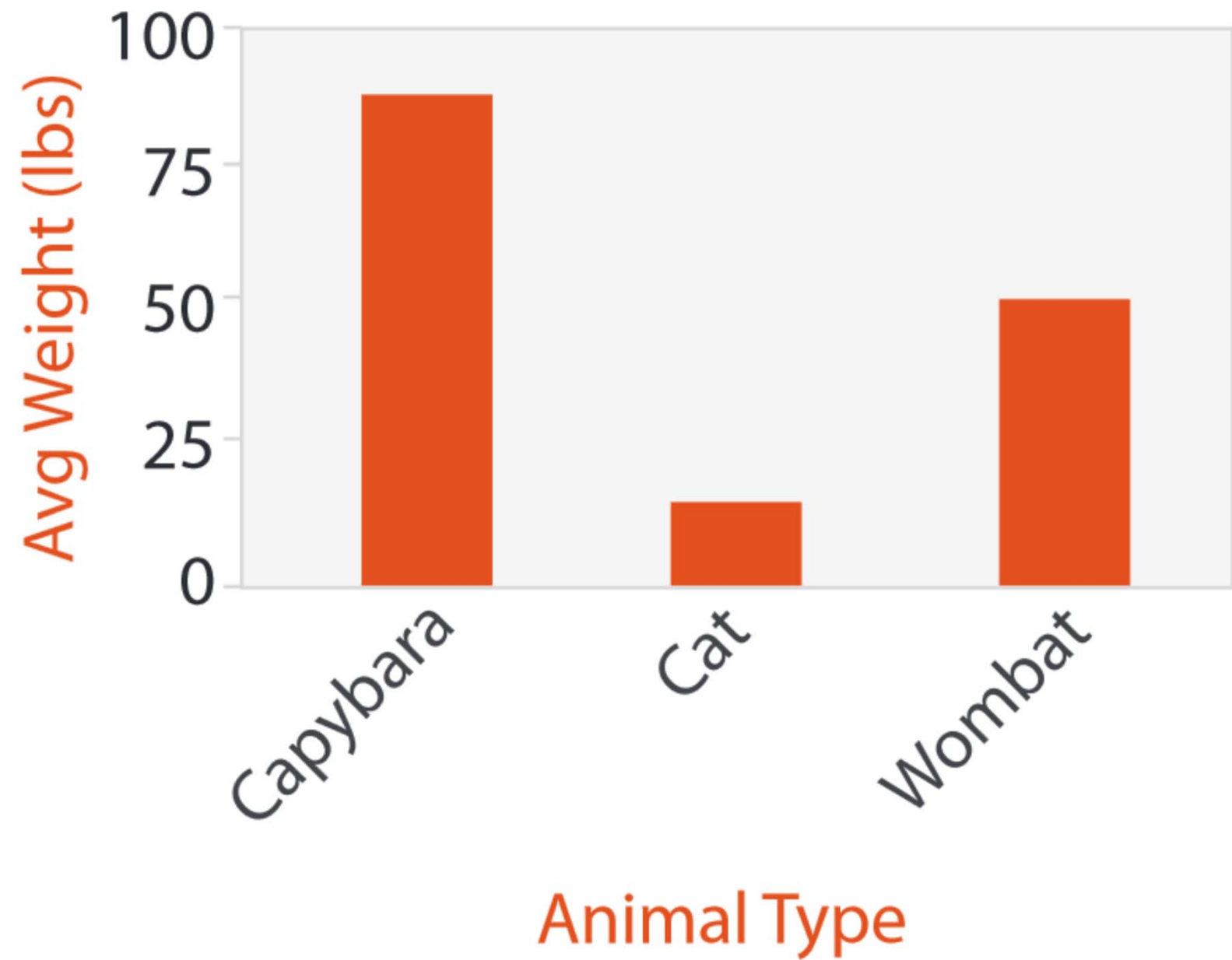
bar charts

ID	Animal type	Avg. weight
1	Capybara	78
2	Cat	13
3	Wombat	50
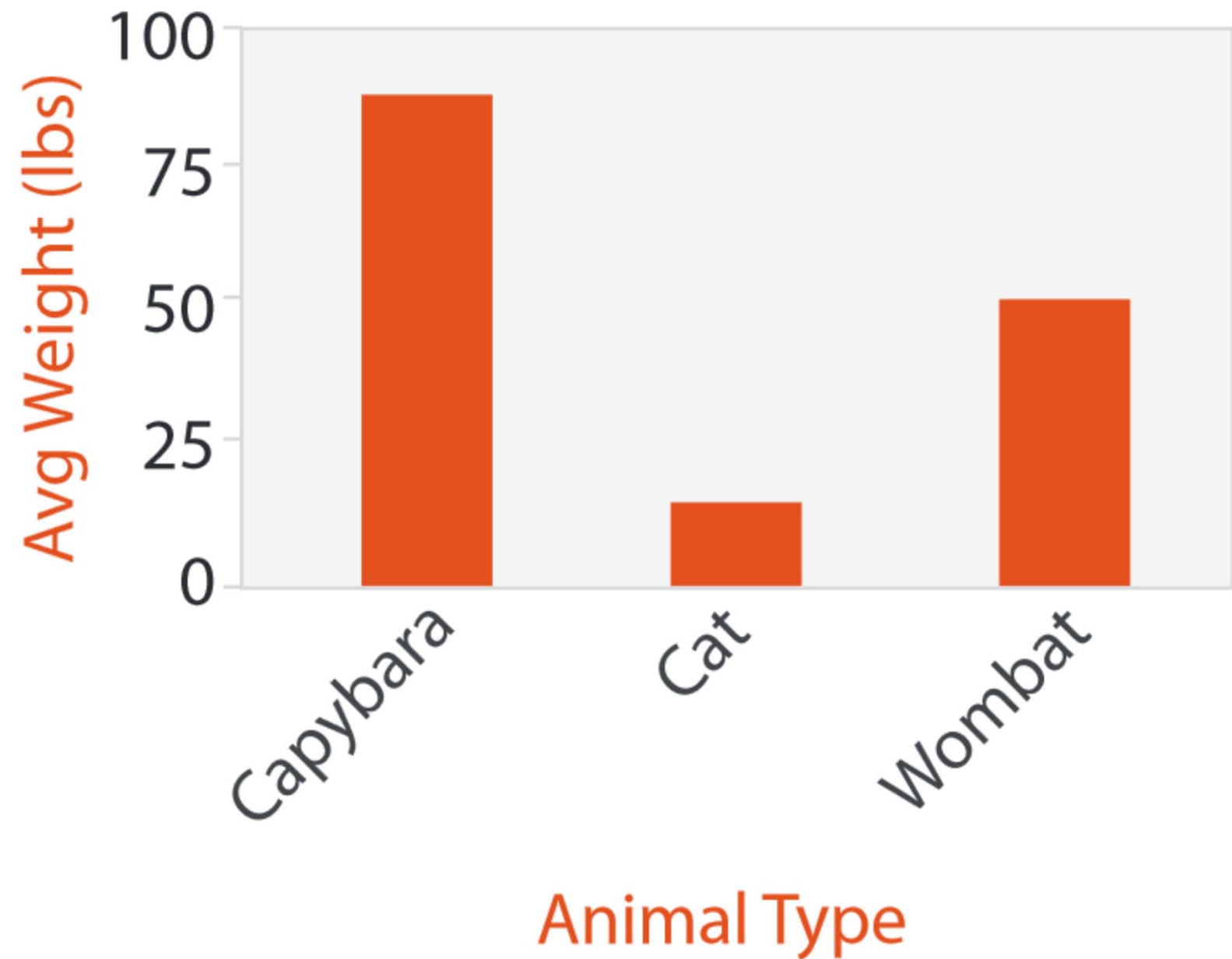
bar charts

One categorical attribute and one quantitative attribute
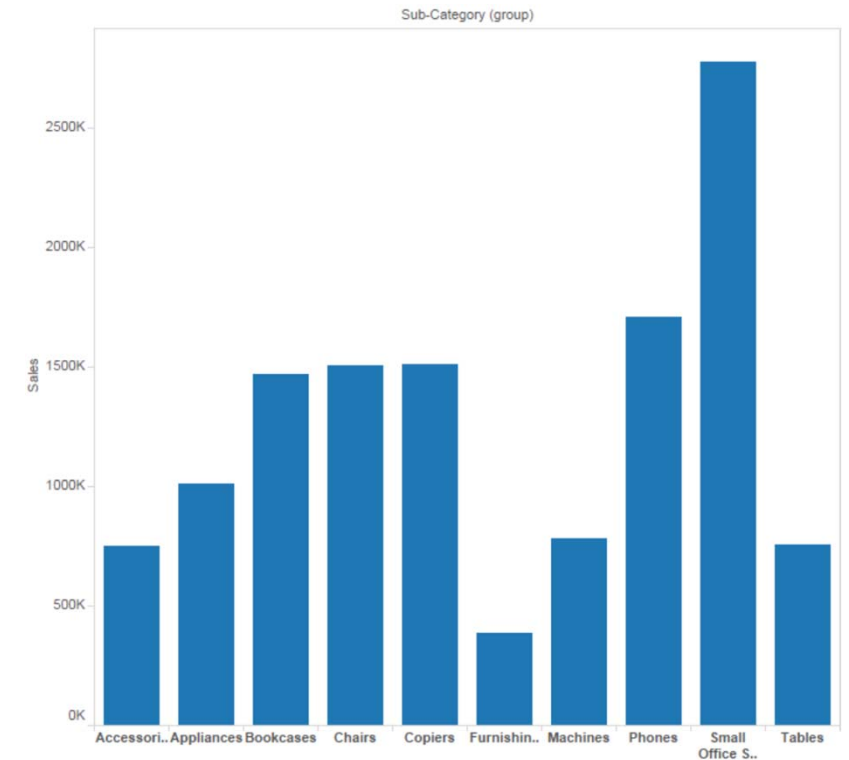
- Line marks

Value attribute expressed through length and aligned in vertical position

- Levels of nominal dimension separated through horizontal position

Lookup

- Compare values

- Presenting trends

bar charts

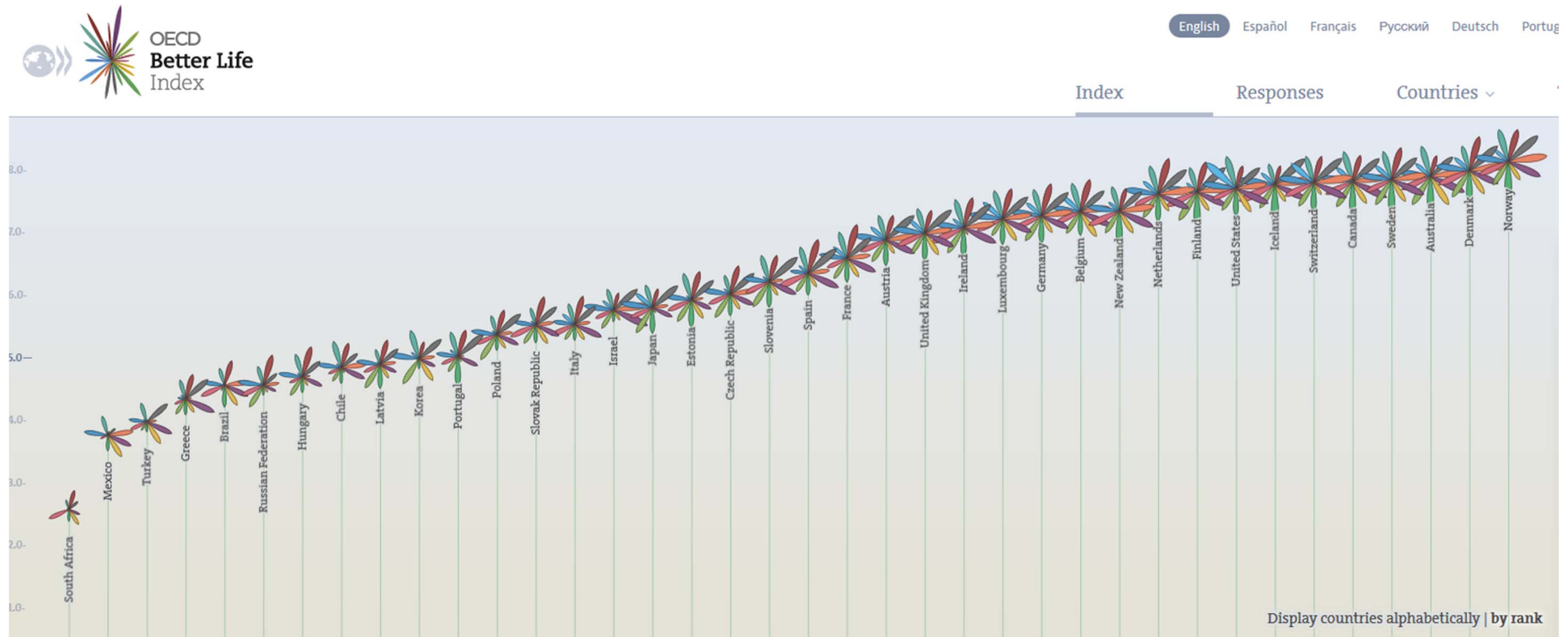http://www.oecdbetterlifeindex.org/#/11111111111

bar charts

- Alphabetical ordering
  - Good for lookup

http://www.oecdbetterlifeindex.org/#/11111111111

bar charts

- Quantity-driven ordering
  - Good for seeing trends

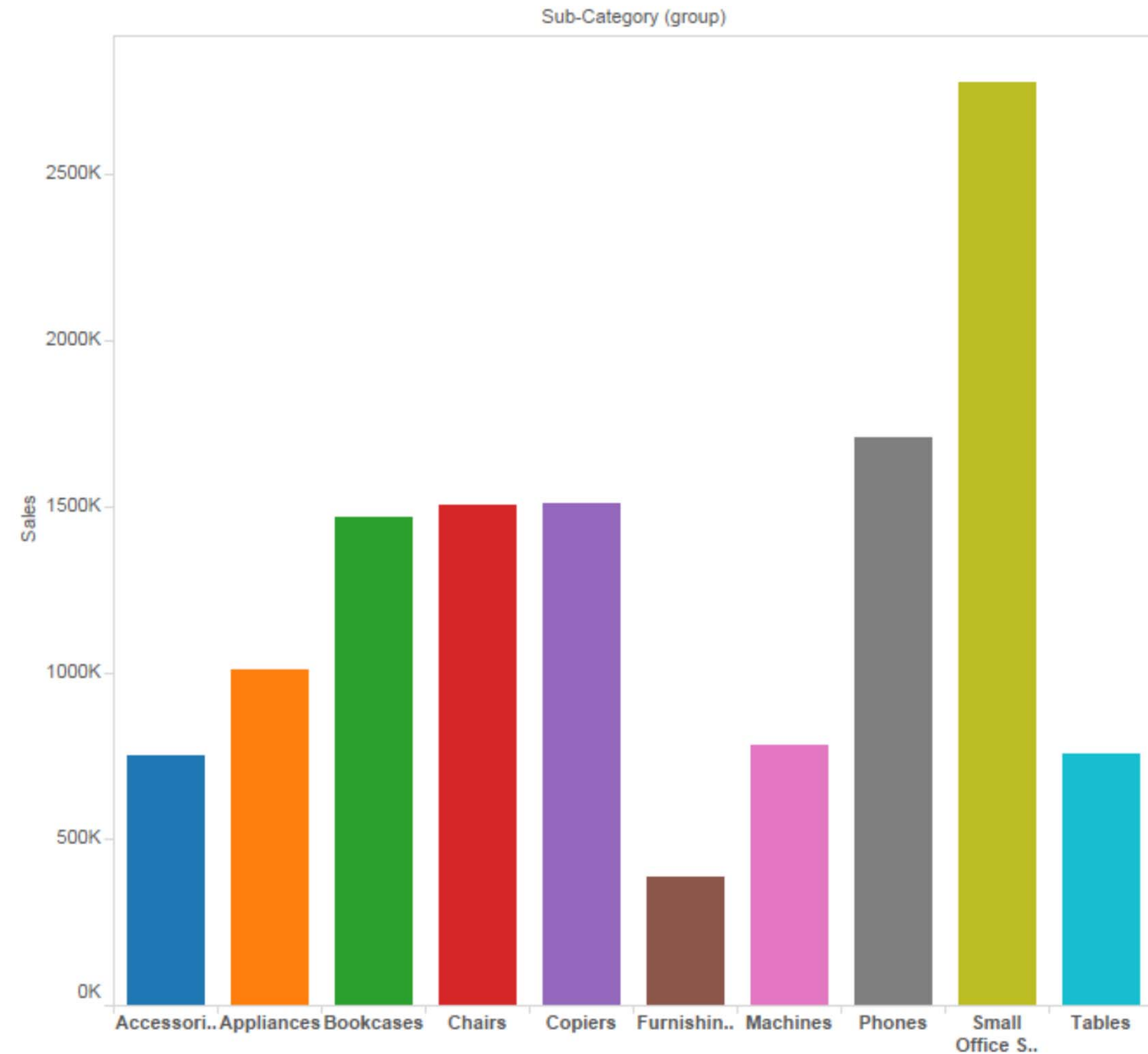http://www.oecdbetterlifeindex.org/#/11111111111
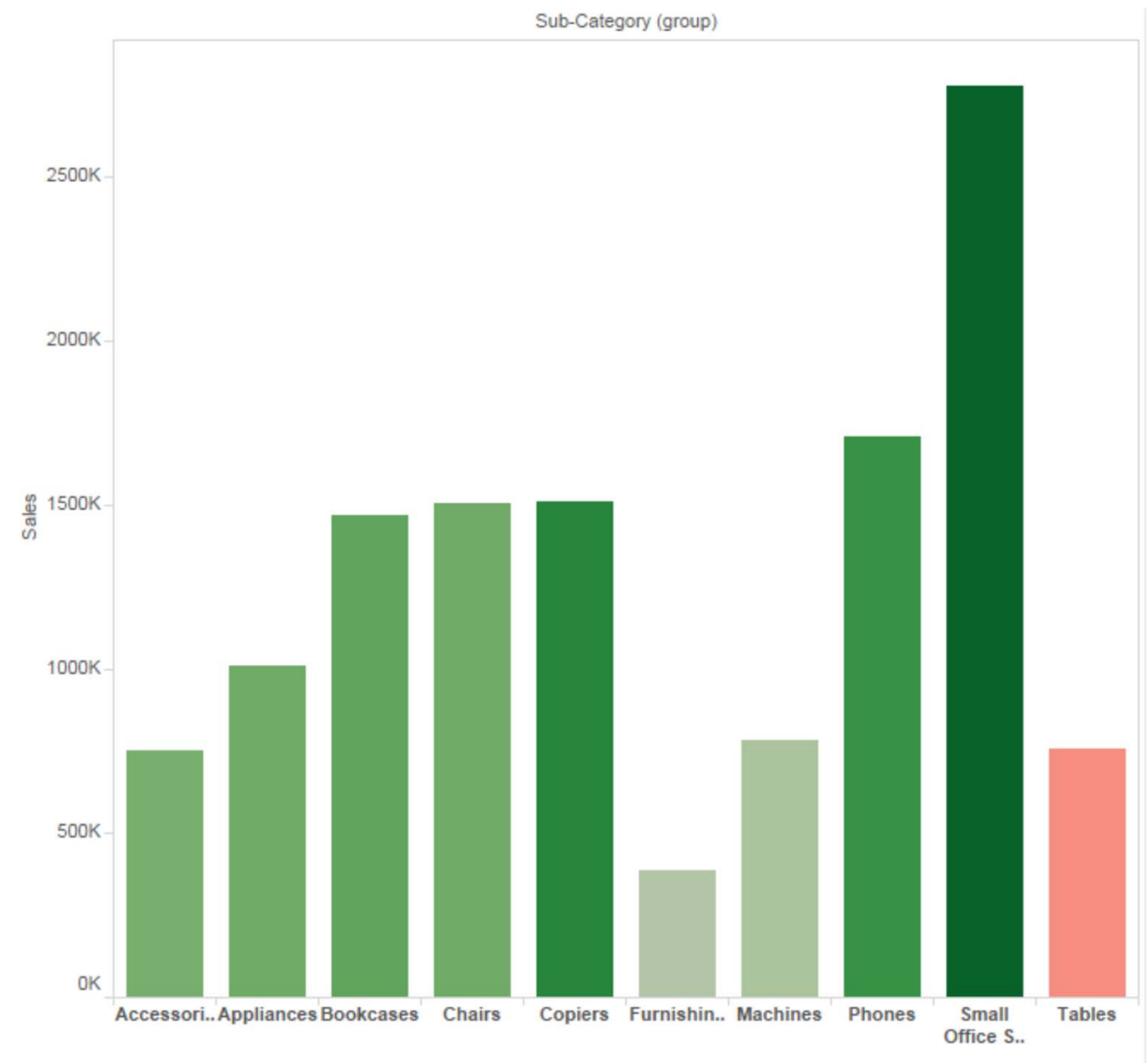
bar charts
double-encoding

- Product category
  - x-position
  - Colour (hue)

- Sales
  - Length

bar charts
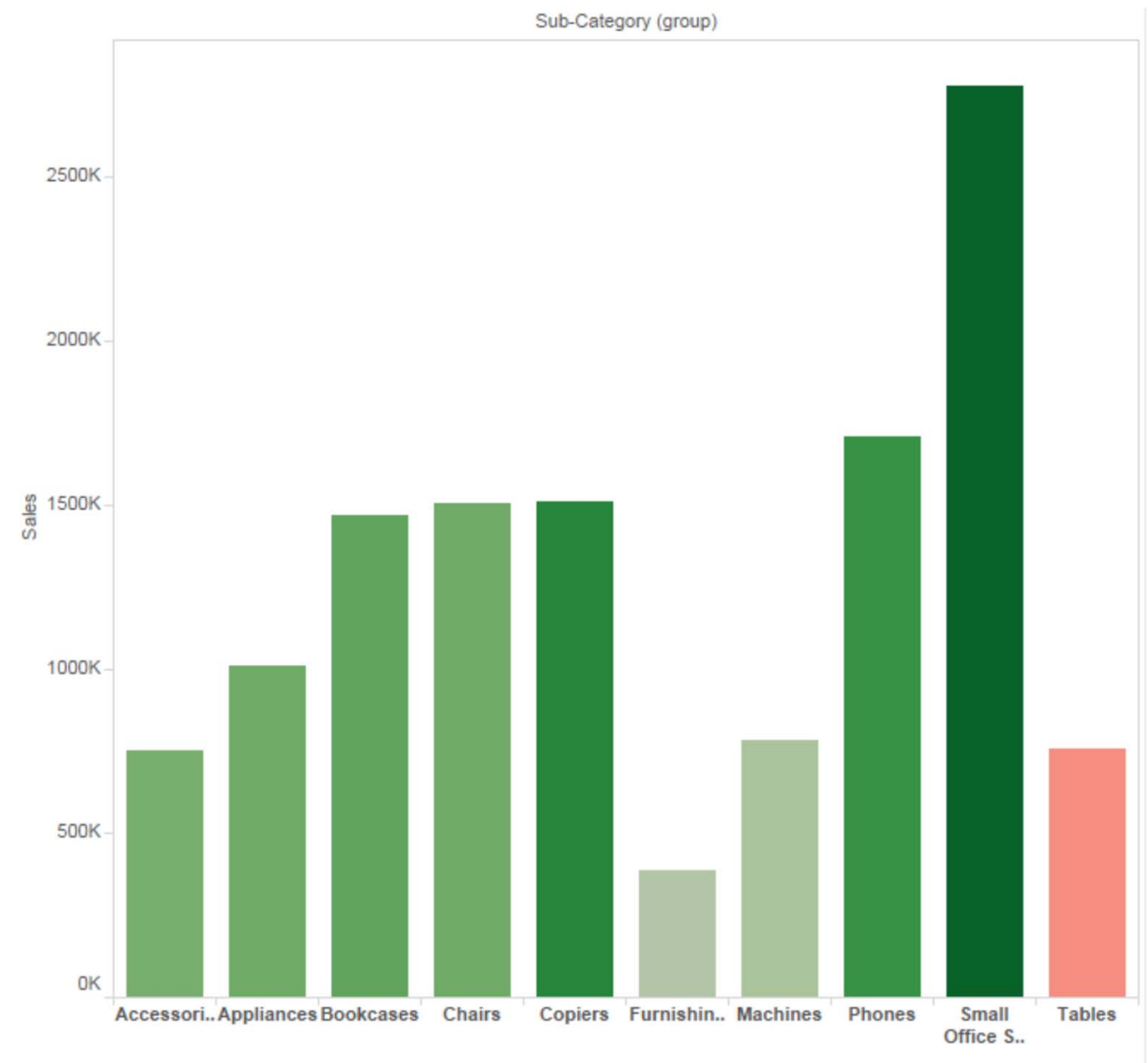one categorical + two quantitative attributes

- Product category
  - x-position
- Sales
  - Length
- Profit amount
  - Colour (saturation)
- Profit category
  - Colour (hue)
  - Red: minus
  - Green: plus
- → Diverging colour scale

Sub-Category (group)

47

bar charts
one categorical + two quantitative attributes

- Product category
  - x-position

- Sales
  - Length

- Profit amount
  - Colour (saturation)

- Profit category
  - Colour (hue)
  - Red: minus
  - Green: plus

→ Diverging colour scale

48

bar charts
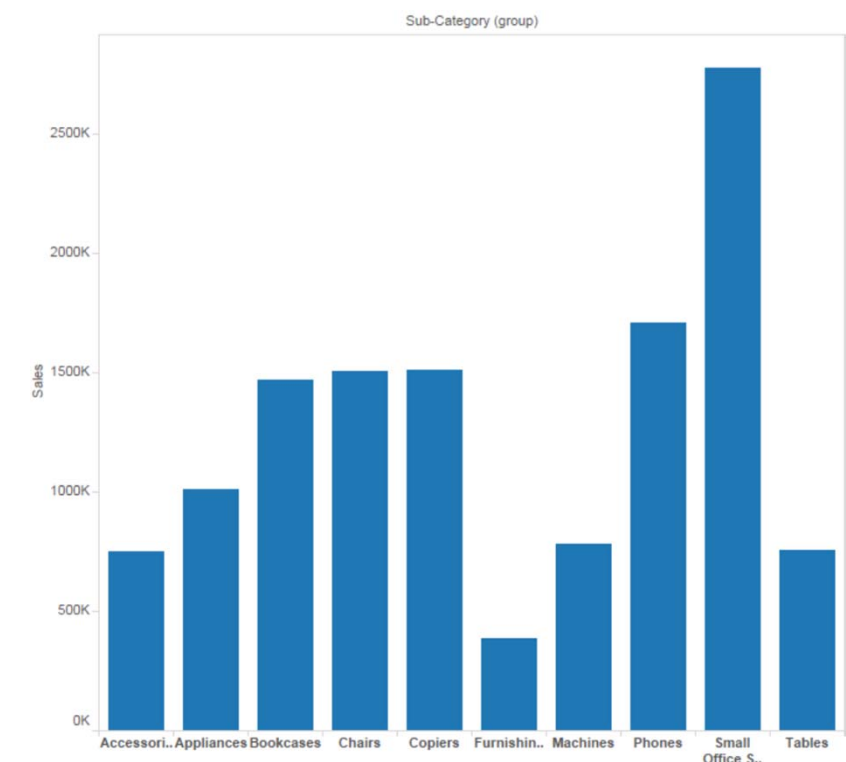
One categorical and one quantitative attribute

- Line marks

Value attribute expressed through length and aligned in vertical position

- Levels of nominal dimension separated through horizontal position

Lookup

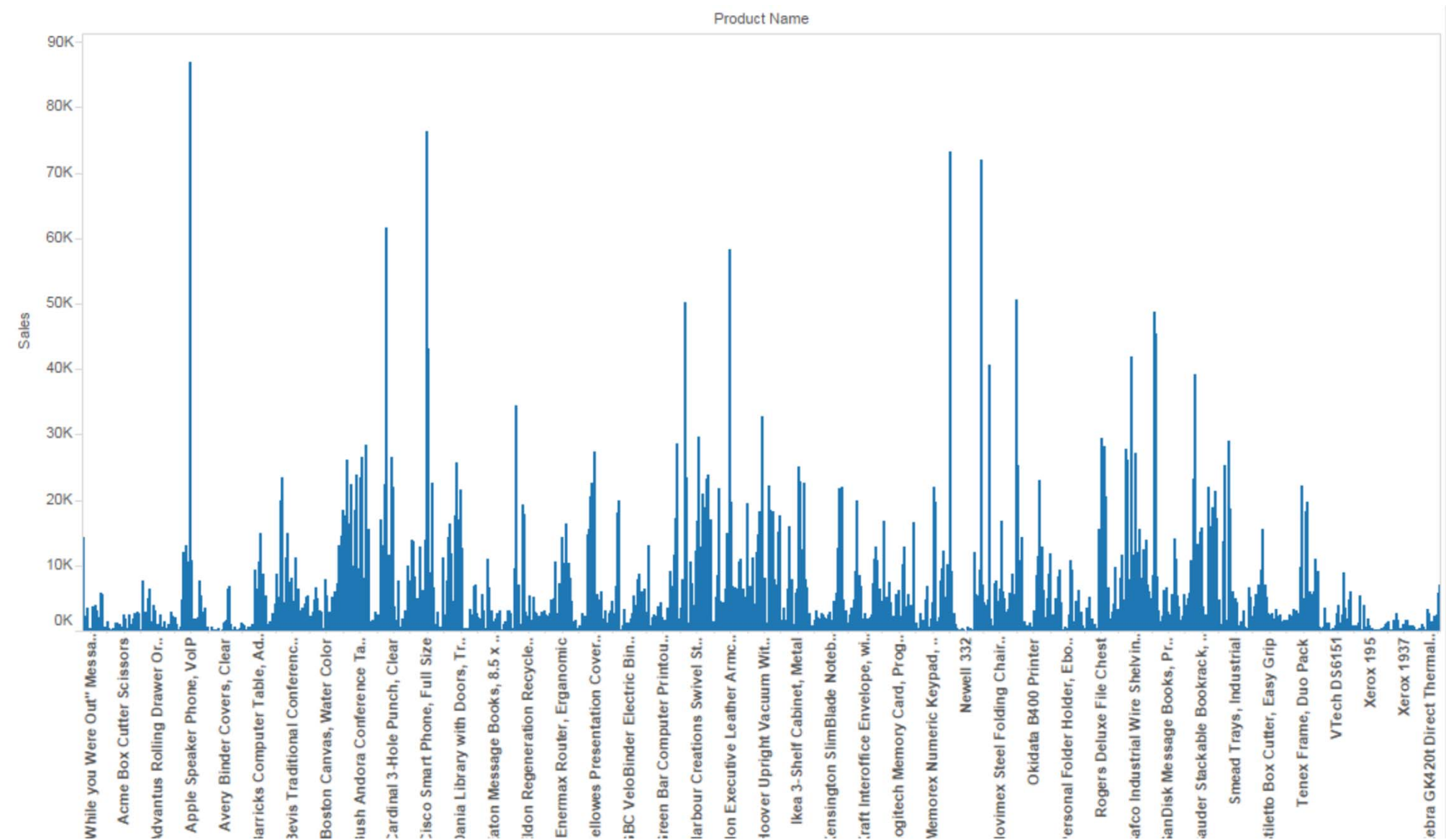- Compare values

- Presenting trends

Limited to hundreds of key levels


49

bar charts

- Over 3000 levels (unique categories)

bar charts & multidimensional data

- Multidimensional table
  - One categorical attribute can be divided into several sub-categories
- Example:
  - Data on different individual customers buys different types of goods:
  - Furniture
  - Office supplies
  - Technology
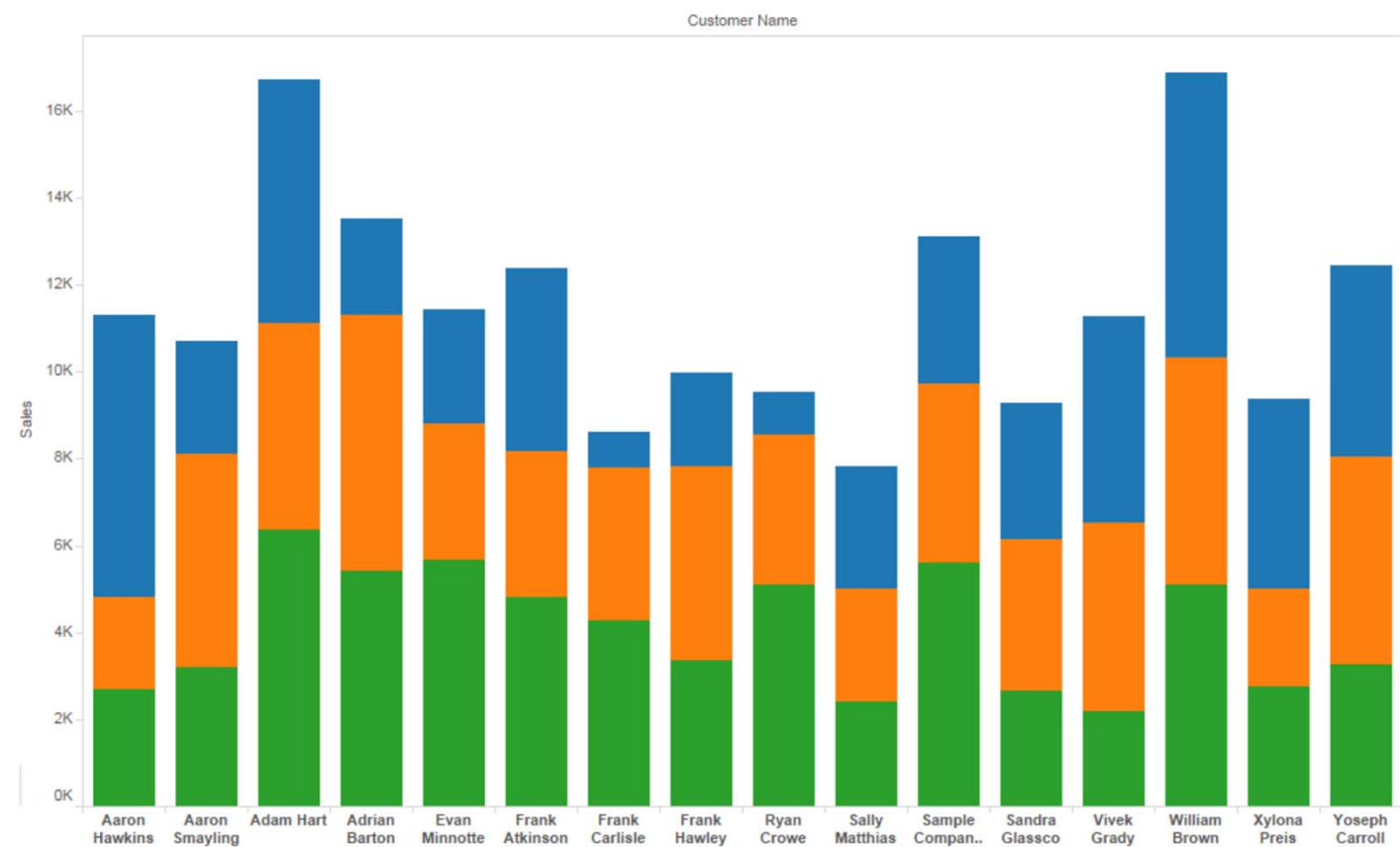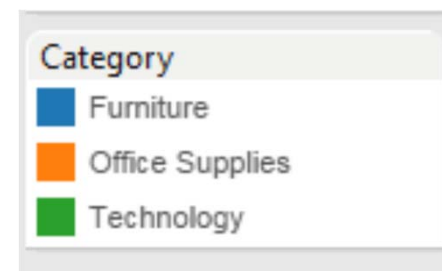 - We want to show the total of sales per customer and category of goods

stacked bar charts
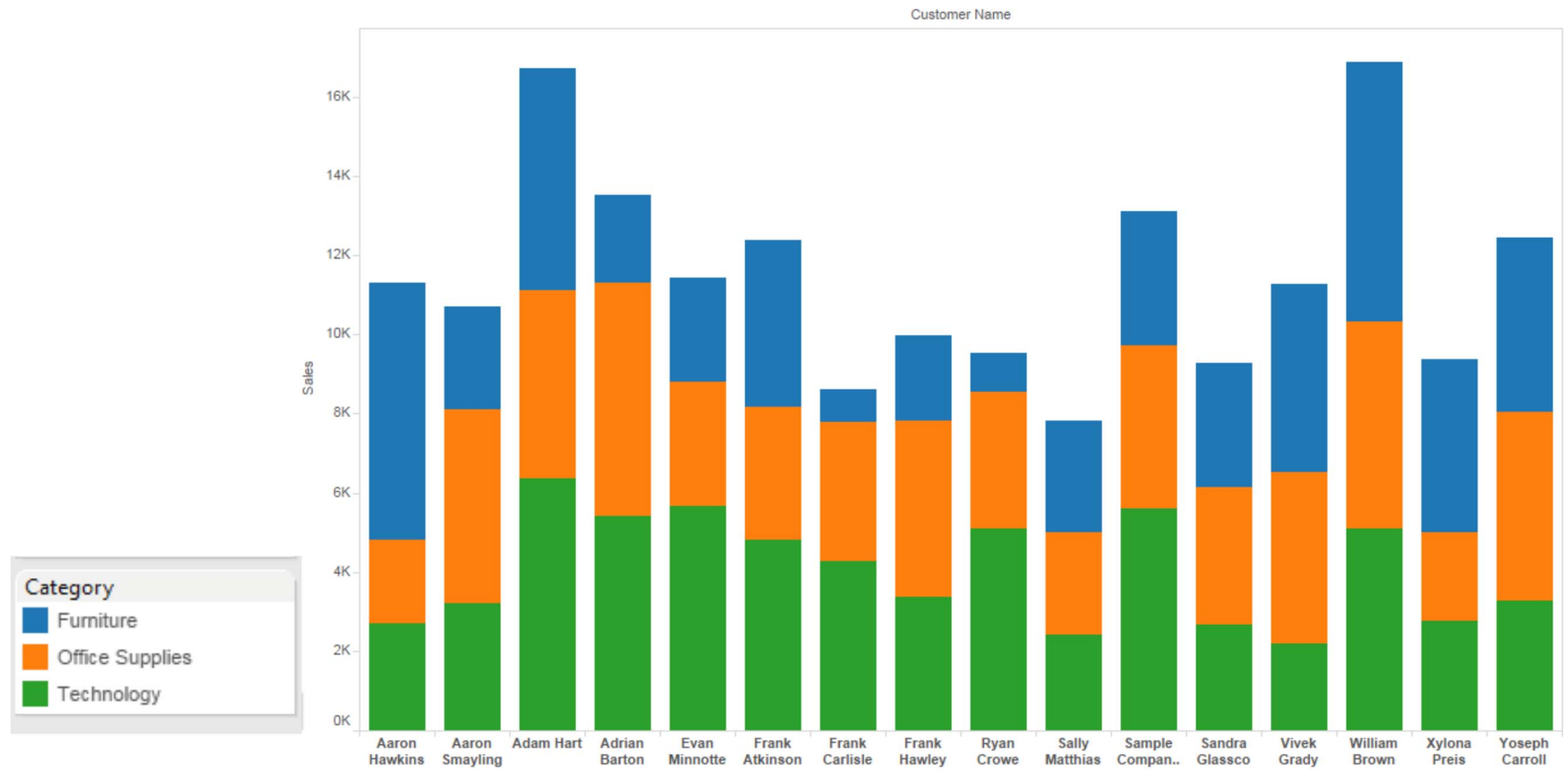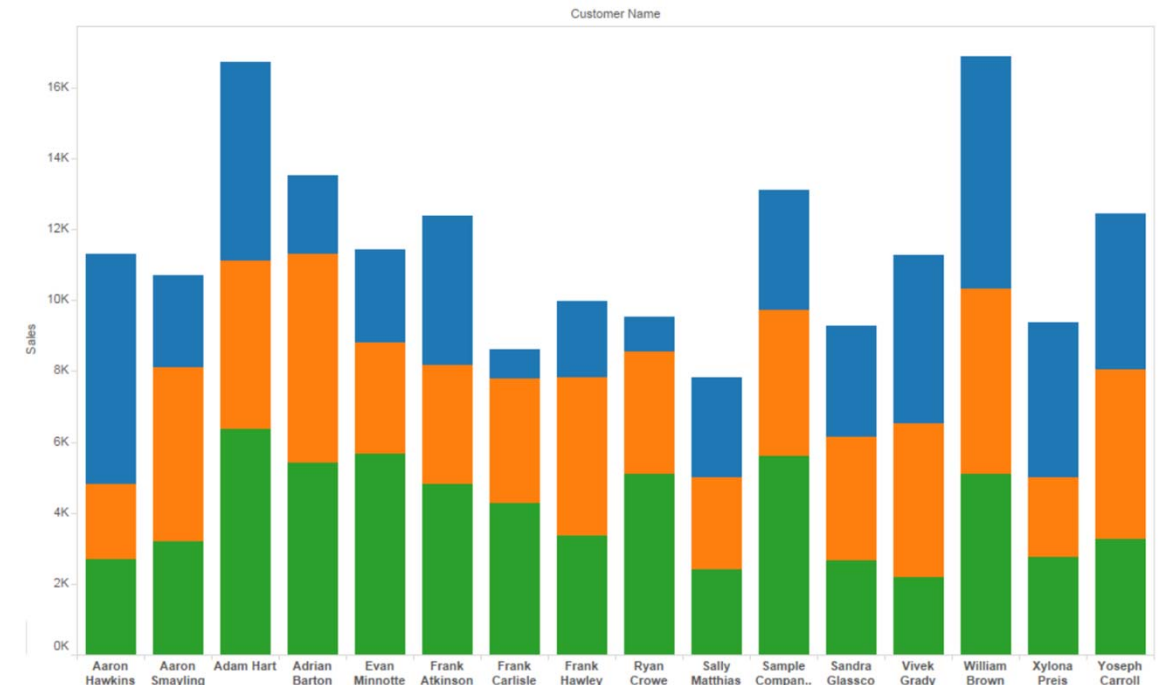
stacked bar charts

what 2 categorical attributes; 1 quantitative attribute

- Customer (primary category)
- Product category (secondary category)
- Sale (quantitative attribute)

stacked bar charts

what 2 categorical attributes; 1 quantitative attribute
- Customer (primary category)
- Product category (secondary category)
- Sale (quantitative attribute)

how Bar glyphs where lengths represents the value of each category of the secondary category

→ Full bar height shows sum of all subcategories

- Primary category levels separated through horizontal position

- Secondary category levels separated through colour (hue)
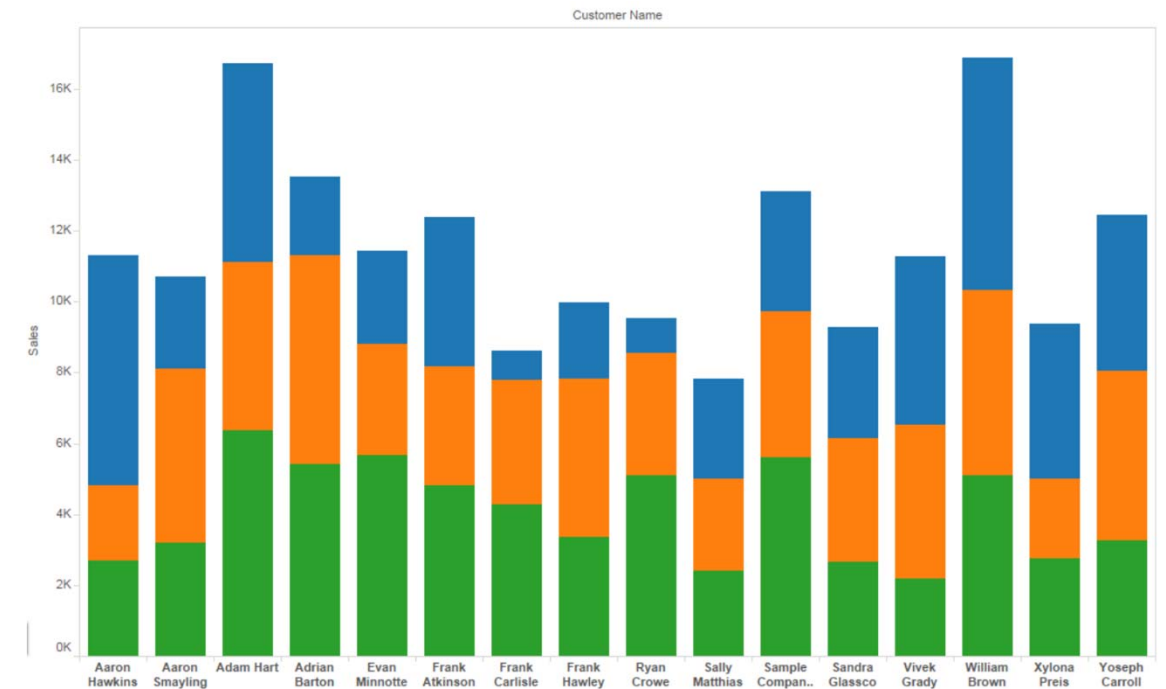
stacked bar charts

why Part-of-a-whole relationship using absolute values

- Lookup values

- Find trends

- Limitations
  - Height of full bars and lowest sub-bars easy to compare
  - Comparison of stacked sub-bars more difficult: no shared baseline
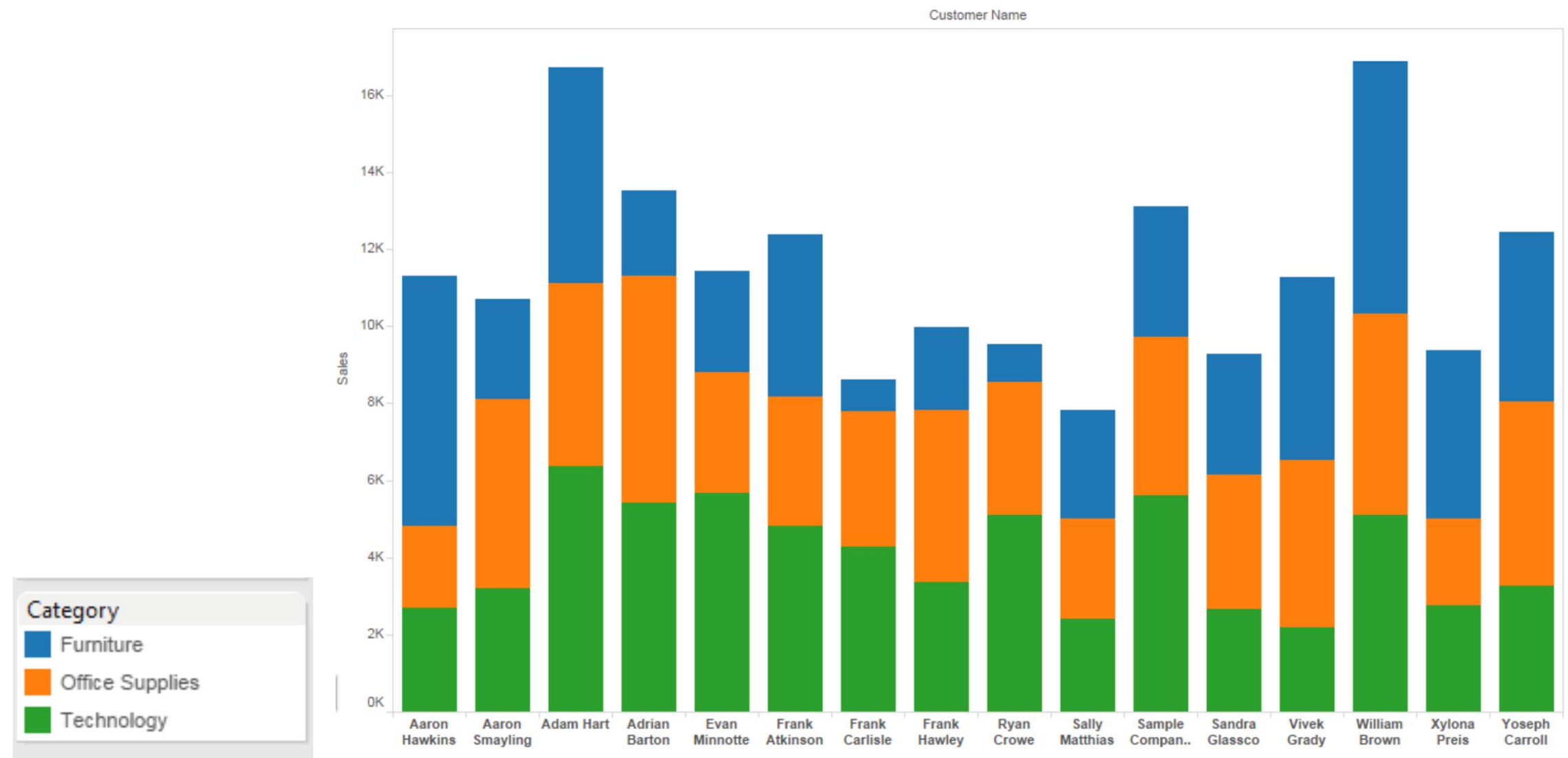 - →Stacking order is important; interactivity can help

scale Scalability:
  - Limited to hundreds of primary category levels
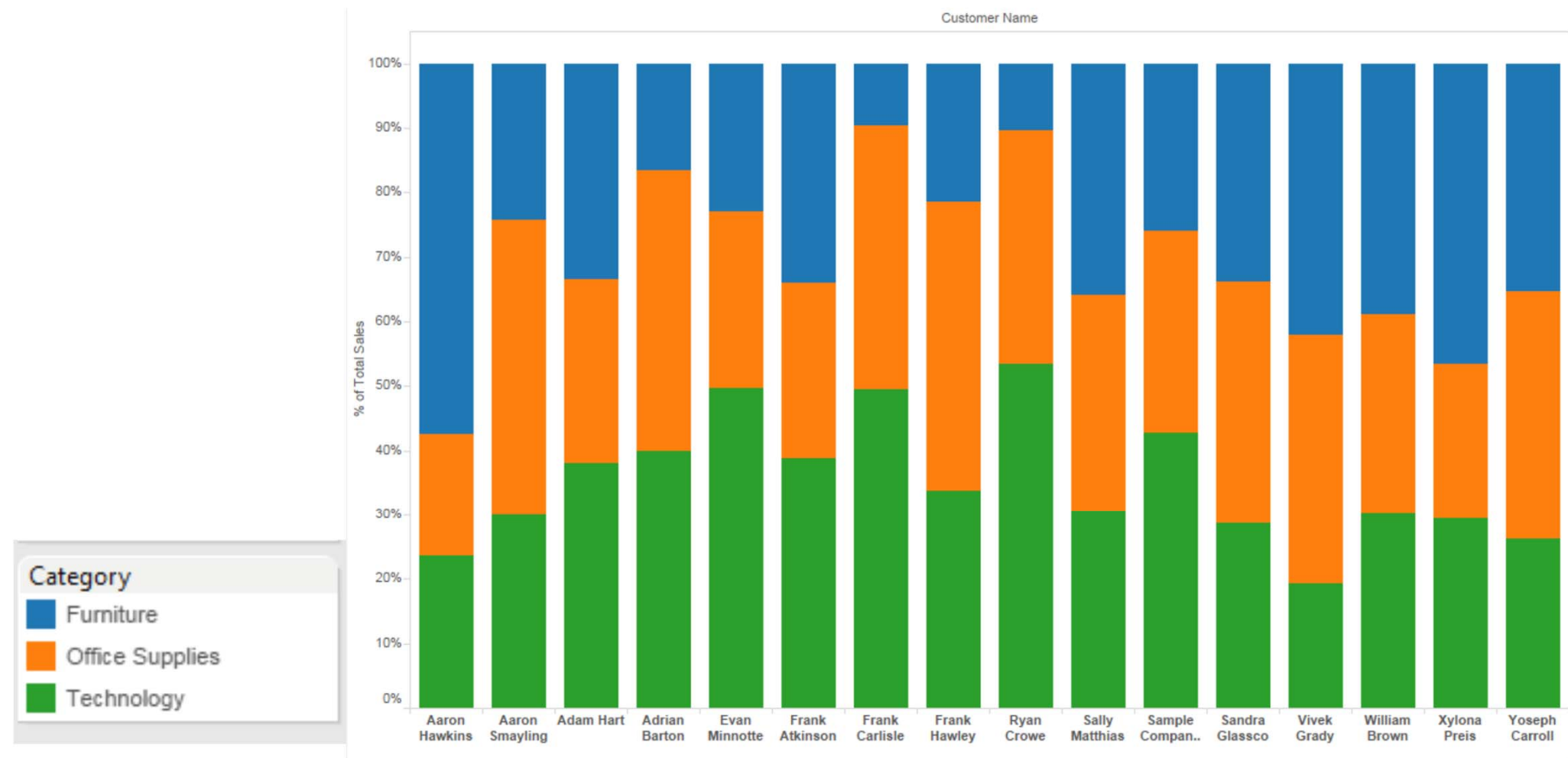 - Limited to several to a dozen secondary category levels

stacked bar graphs
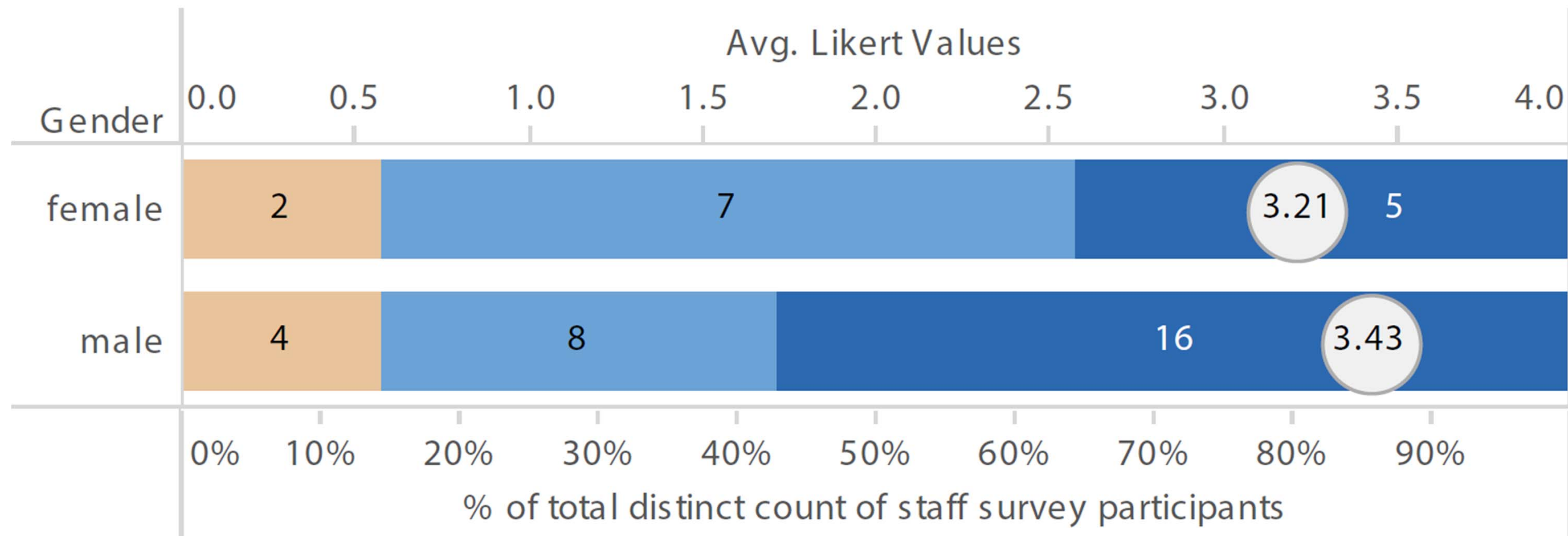
- Typical: absolute values

normalized stacked bar charts

- Showing relative proportions
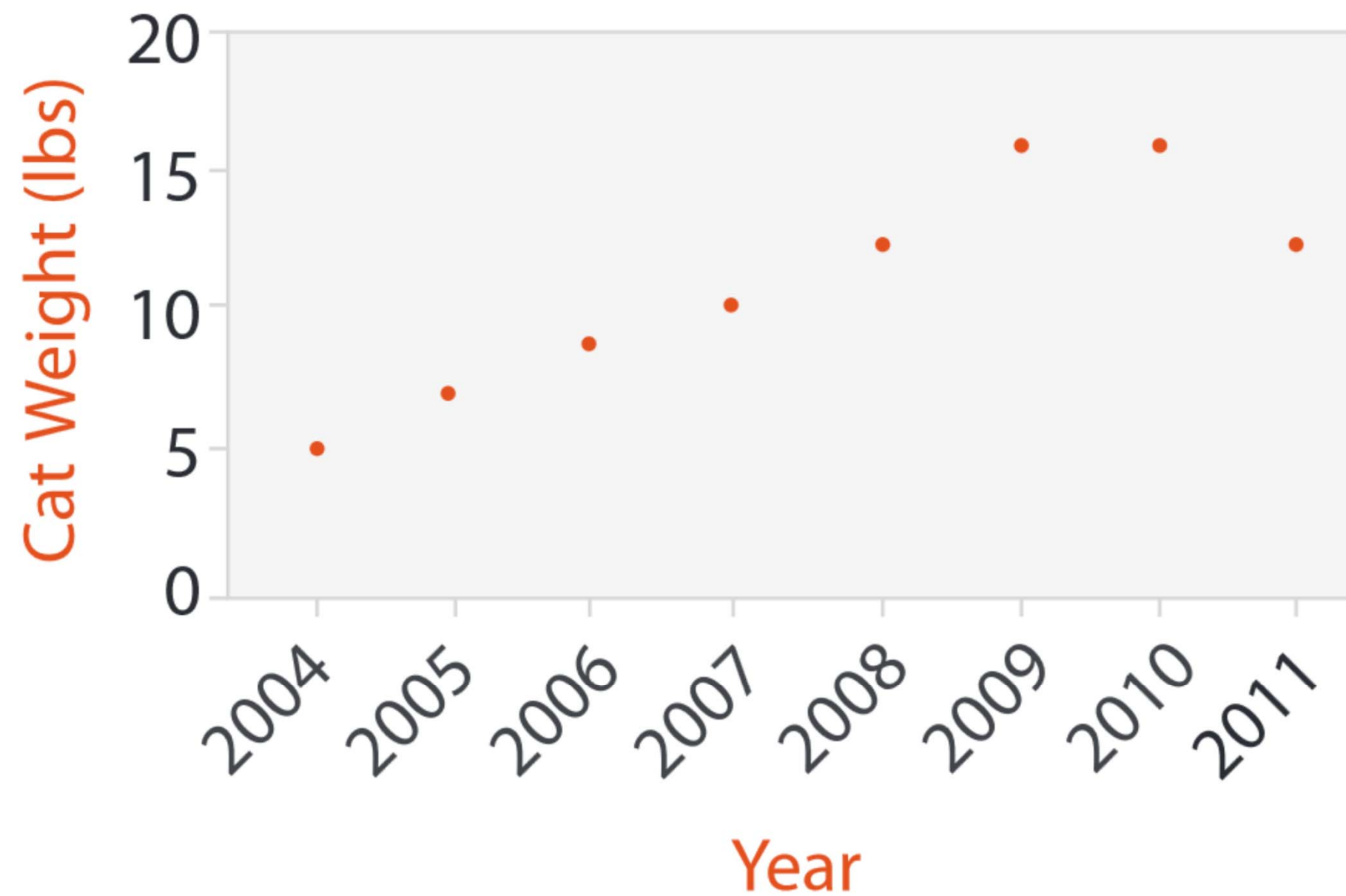- Part-of-whole relationships; comparison of distribution across key attribute levels

normalized stacked bar charts

- Showing relative proportions
- Part-of-whole relationships; comparison across category attribute levels
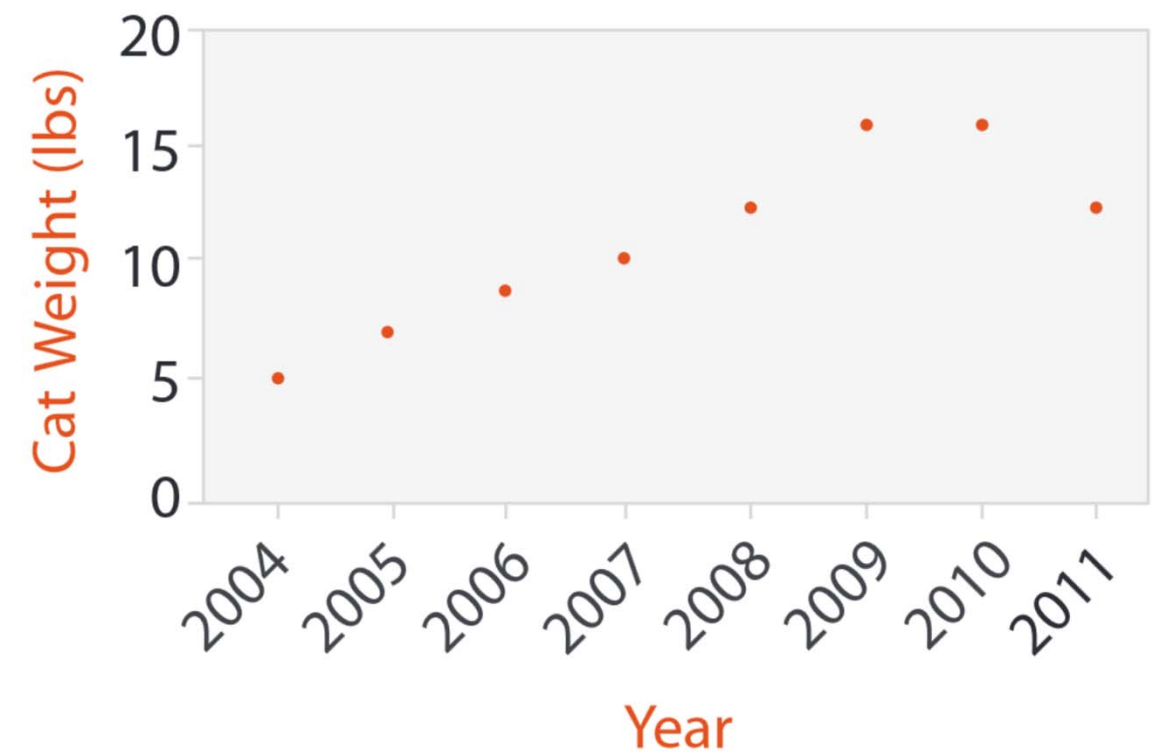
dot chart / dot plot

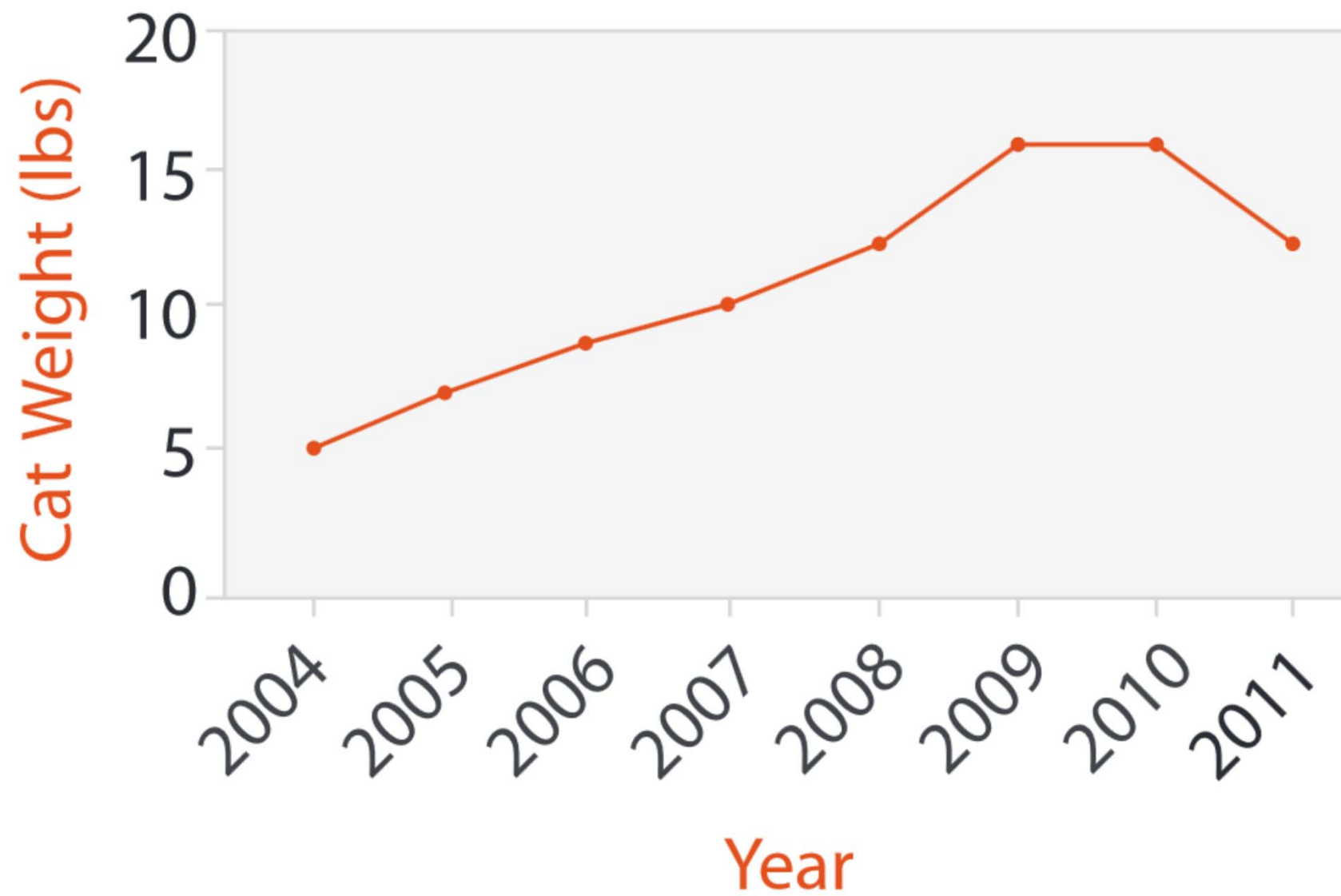# dot chart / dot plot

what Difference to scatter plots

- One quantitative and one ordinal attribute instead of two quantitative attributes

how Difference to bar charts

- Point mark instead of line mark
- Ordinal instead of categorical
- → Dots can imply connection and/or order
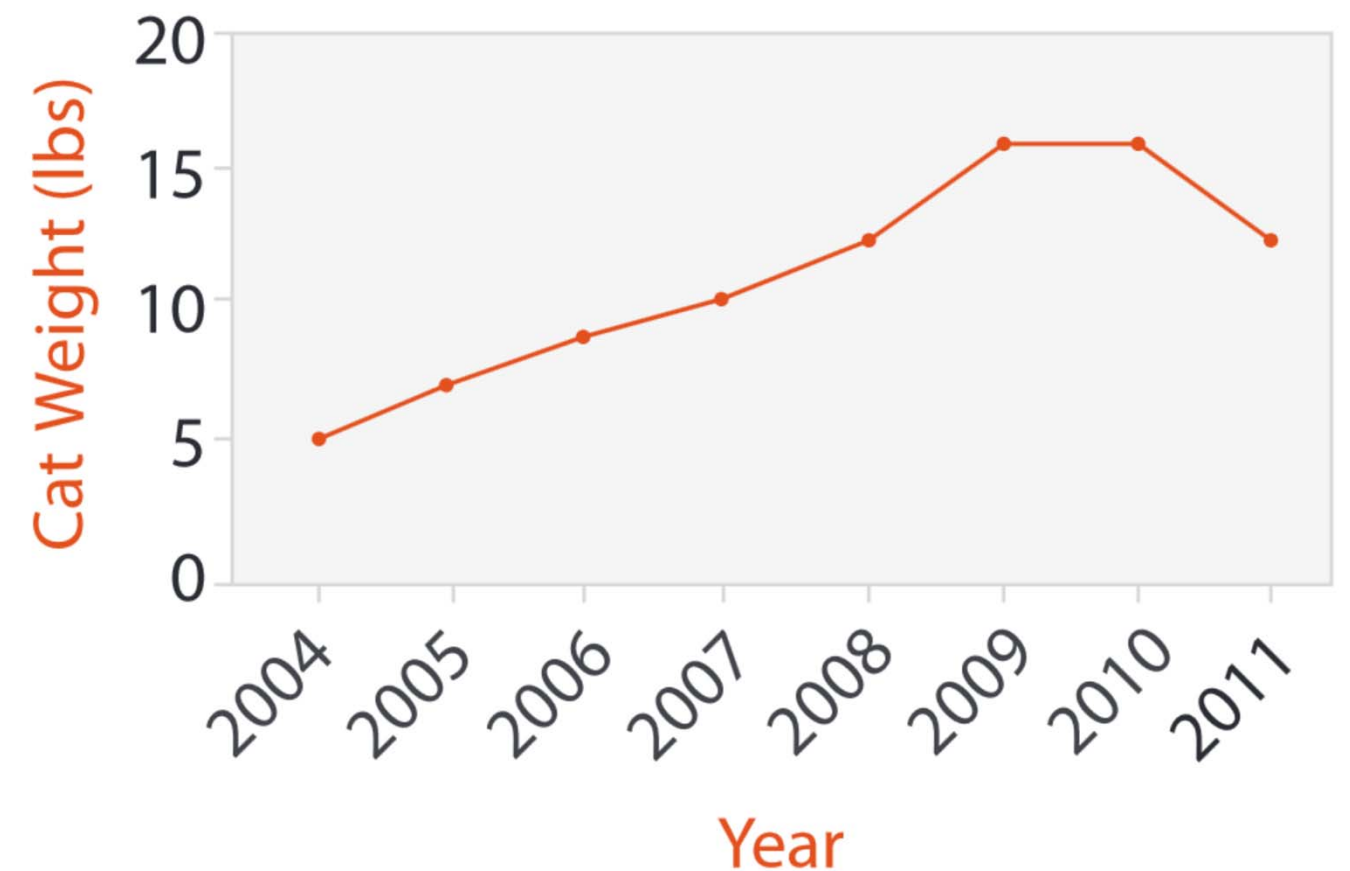
line chart / line graph

line chart / line graph

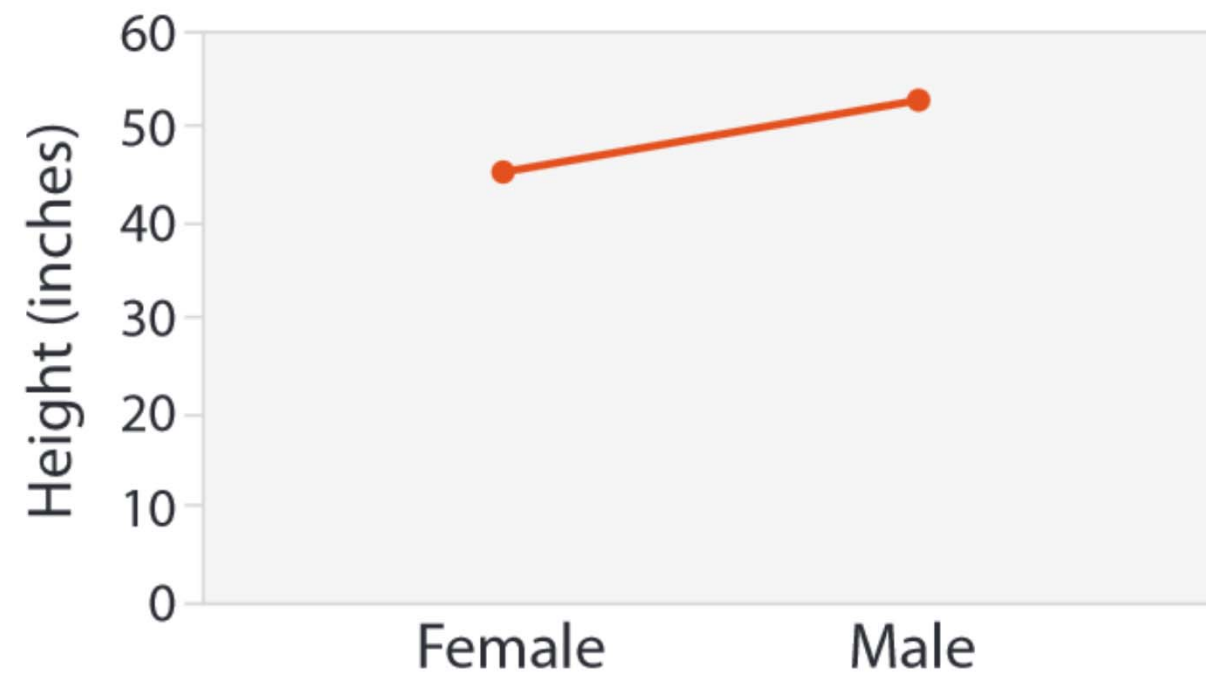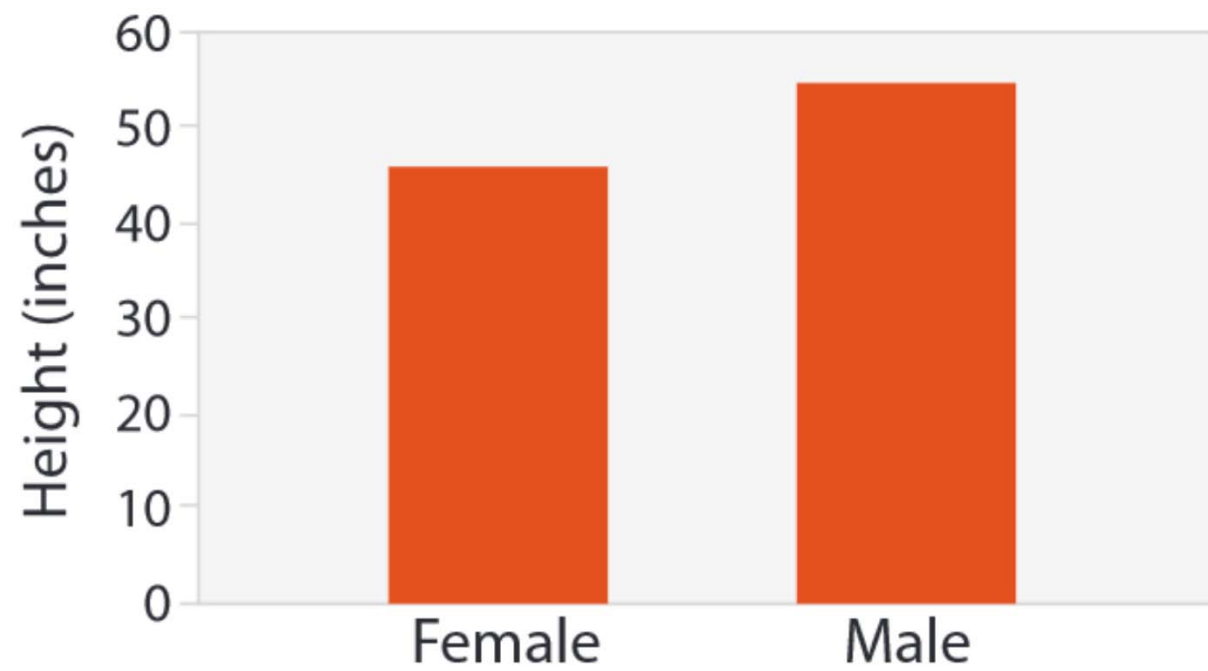- Follow directly from dot charts with a line connecting the different data points

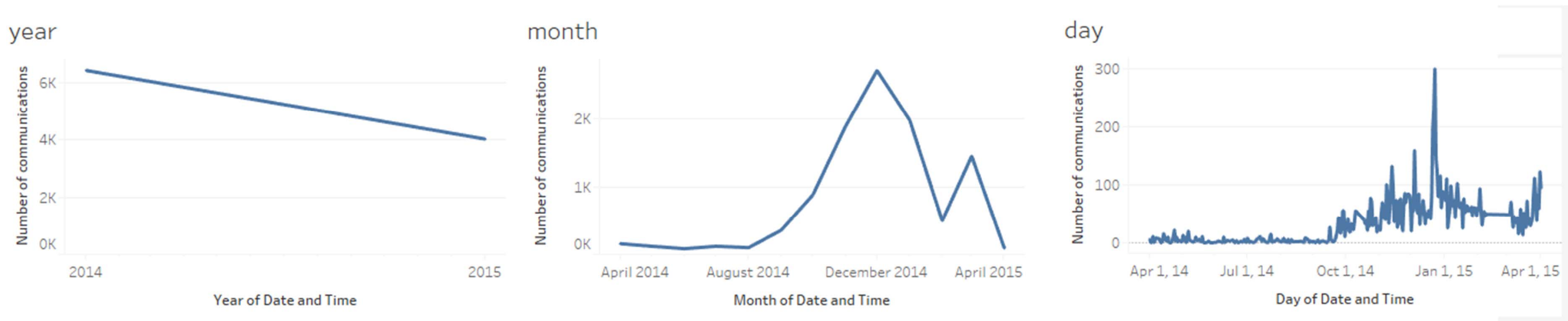why Good for showing trends

line chart / line graph

- Line emphasizes ordering & continuity
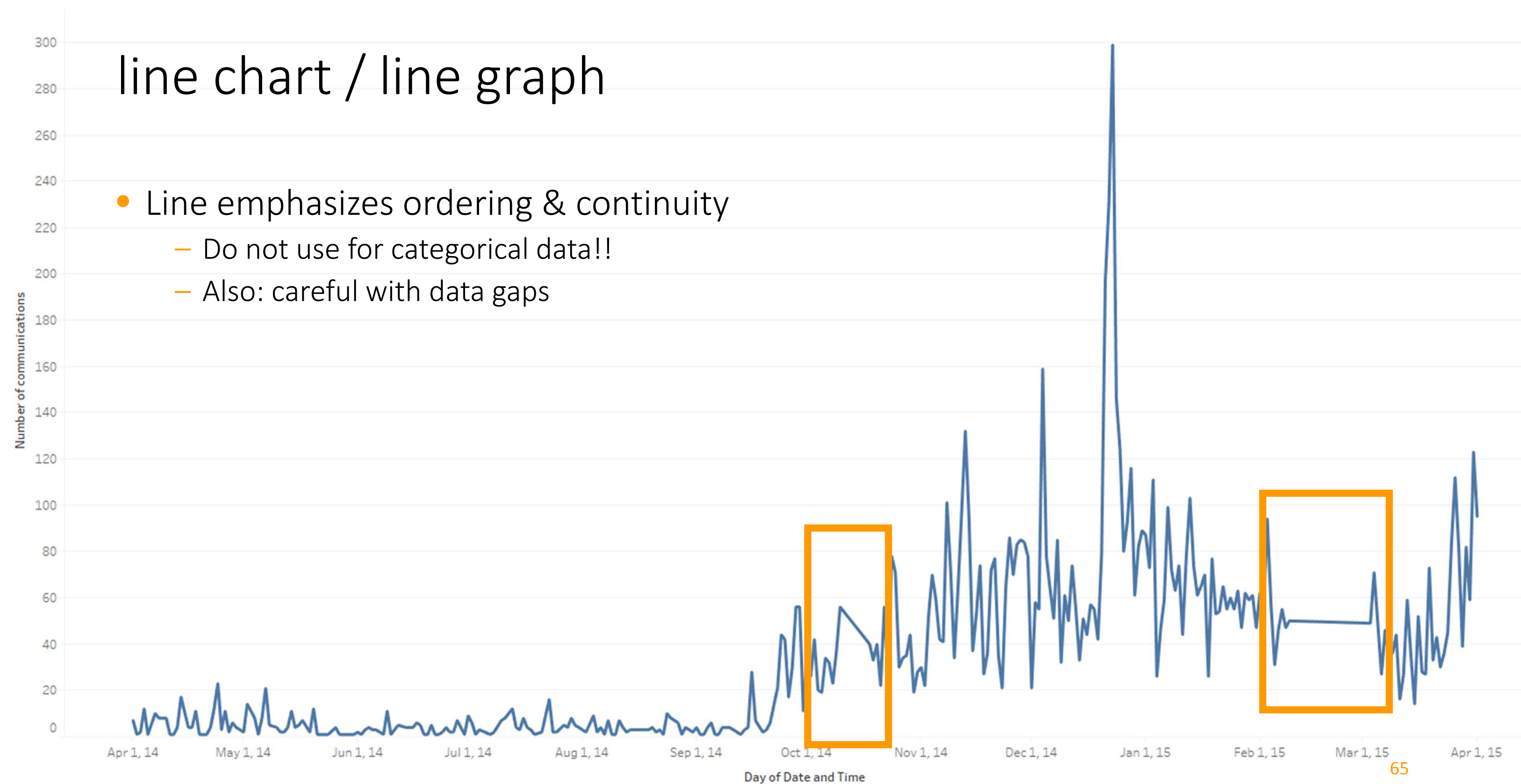 - Do not use for categorical data!!

line chart / line graph

- Line emphasizes ordering & continuity
  - Do not use for categorical data!!
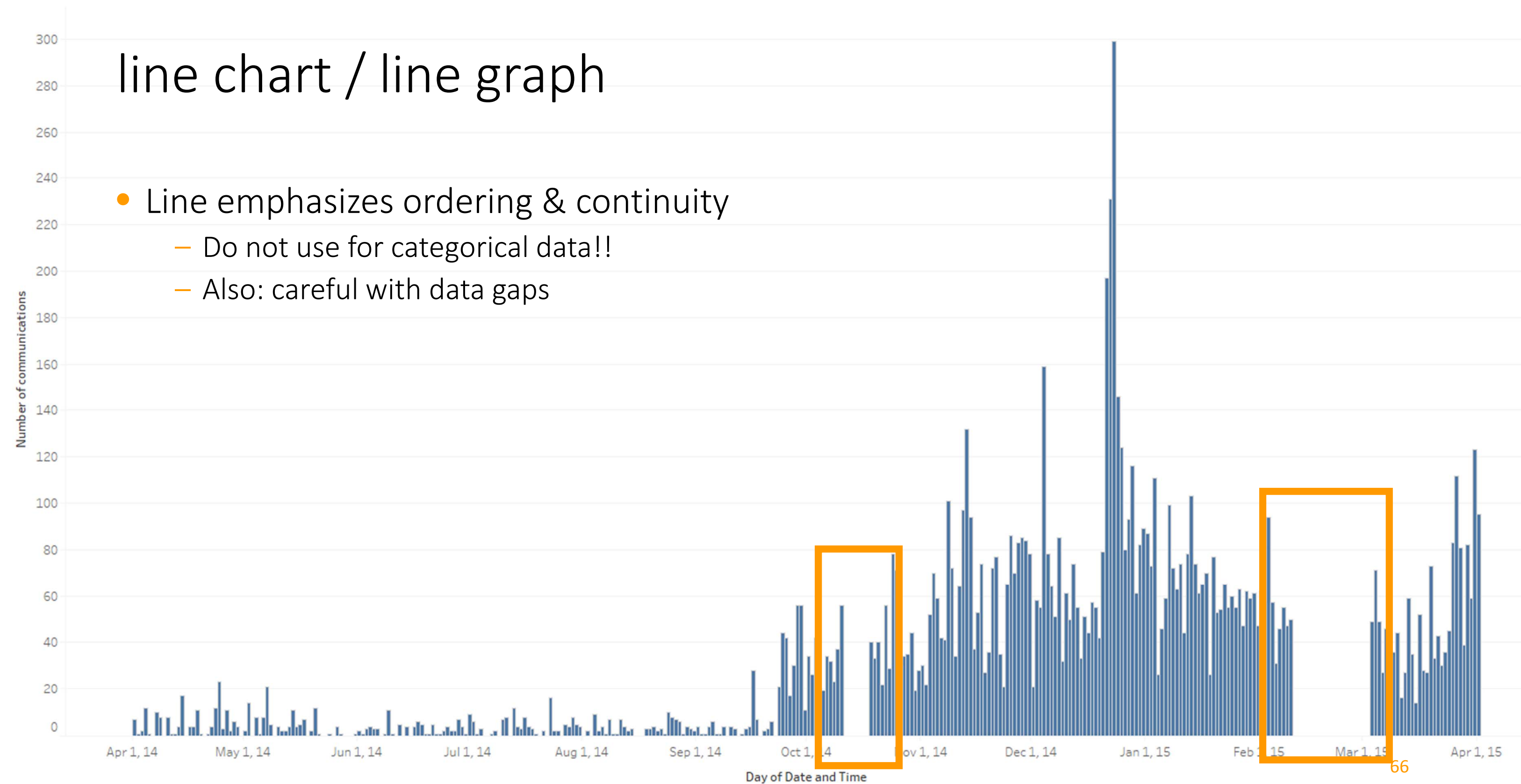 - Also: careful with data gaps

line chart / line graph

- Line emphasizes ordering & continuity
  - Do not use for categorical data!!
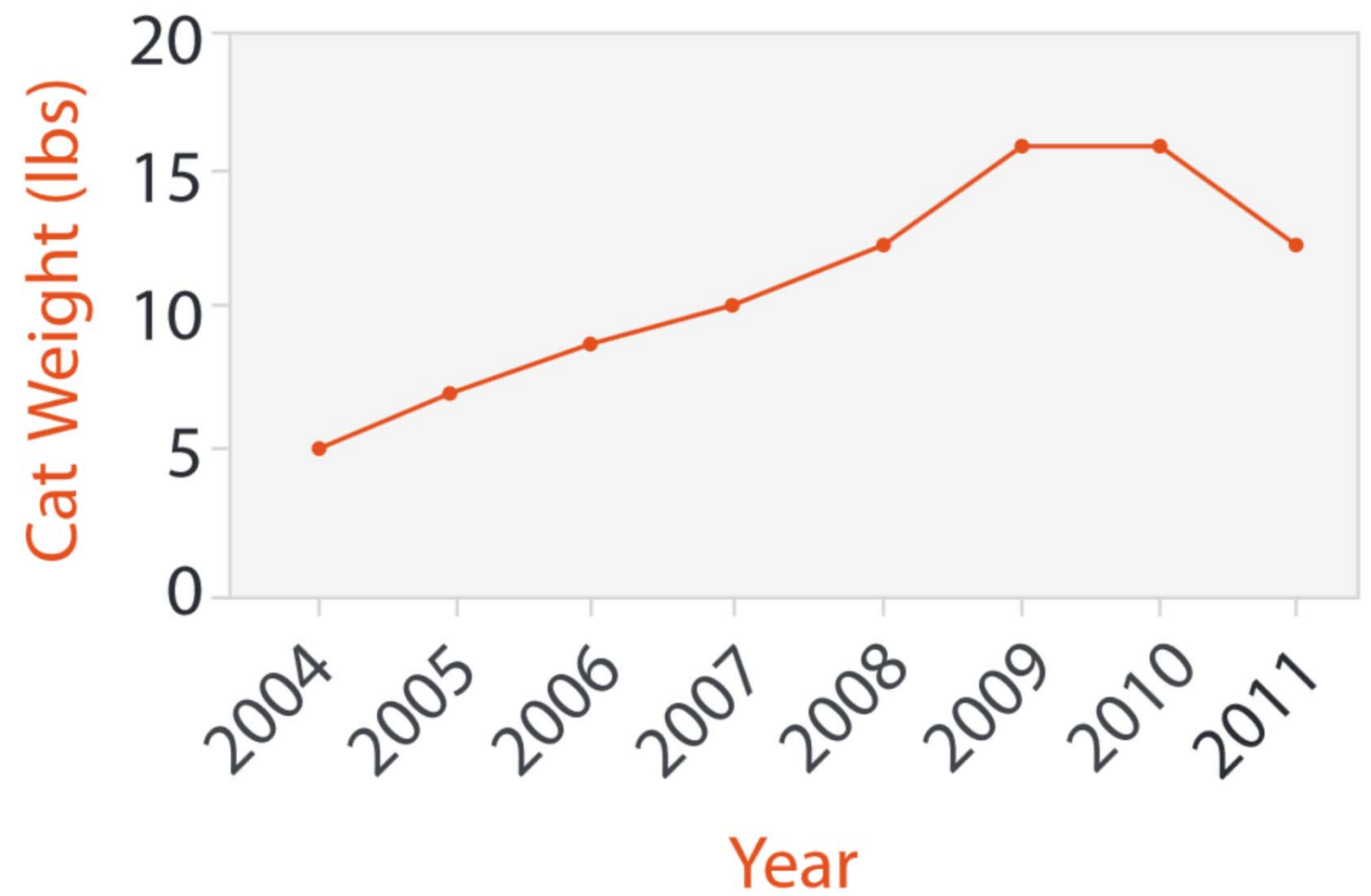 - Also: careful with data gaps

line chart / line graph

- Line emphasizes ordering & continuity
  - Do not use for categorical data!!
  - Also: careful with data gaps

line chart / line graph

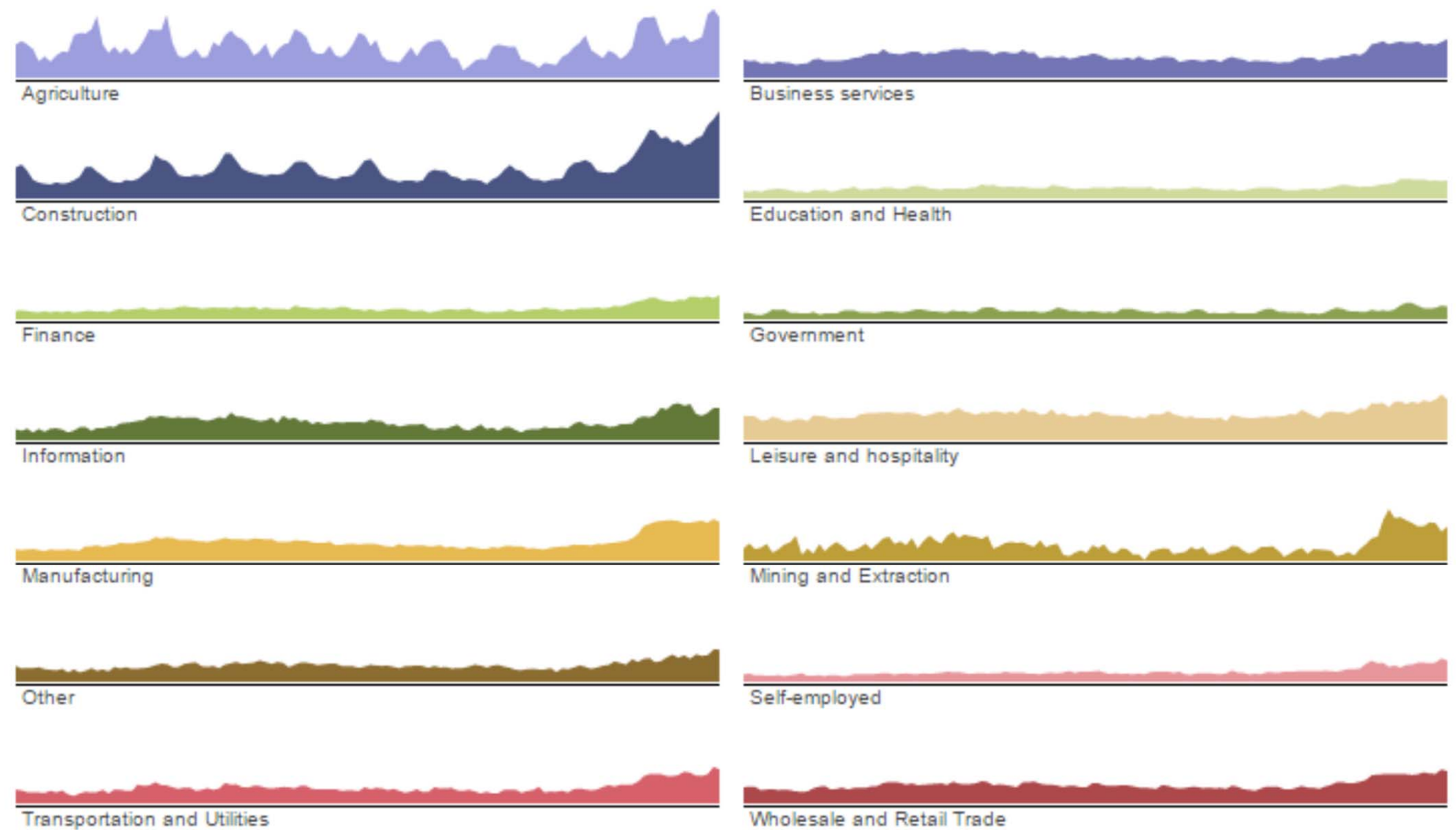Primary attribute (ordinal): hundreds of levels

showing multiple attributes through line charts

- Primary attribute: year (ordinal)
- Secondary attribute: product category (categorical)
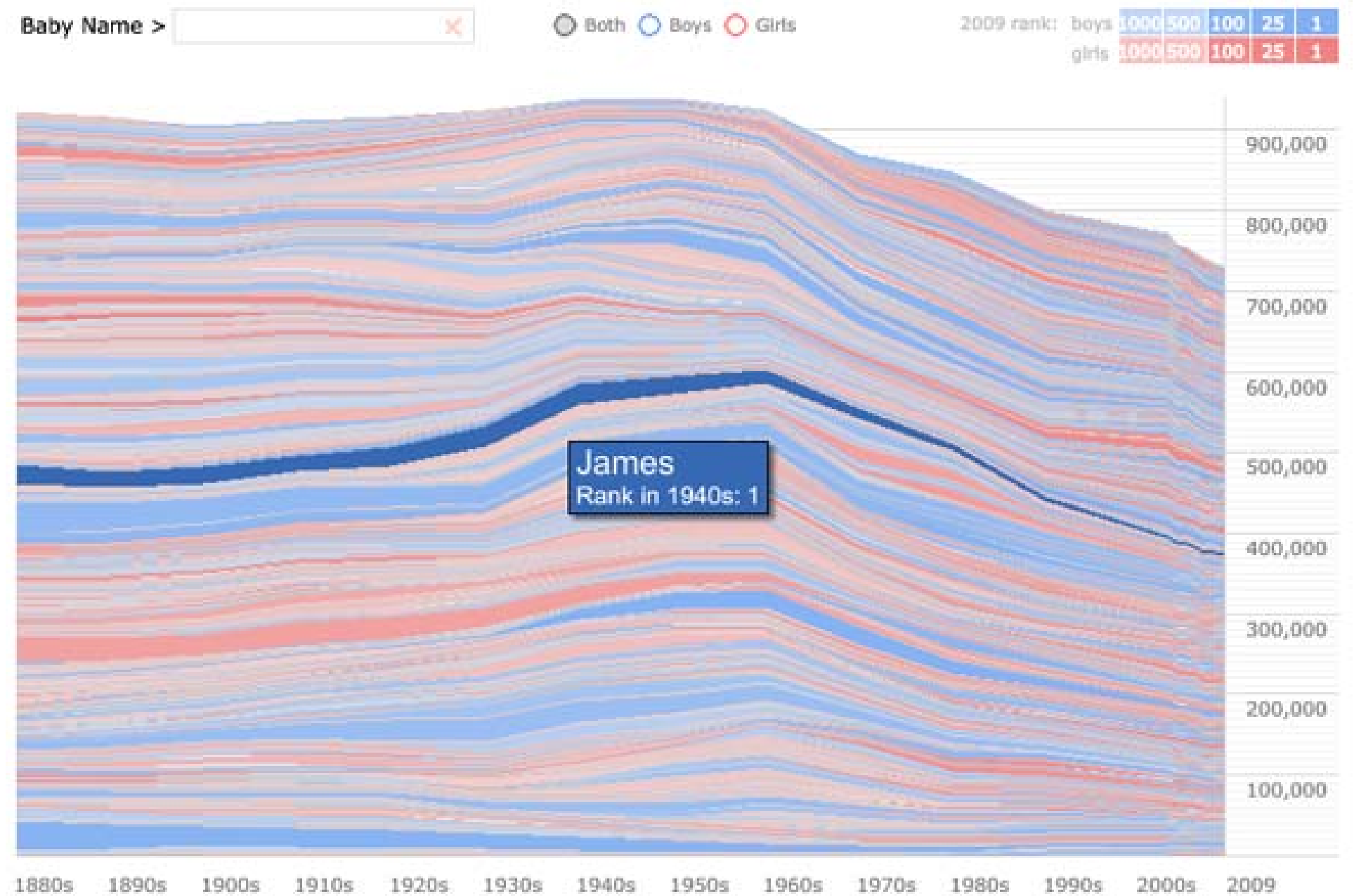- Quantitative attribute: profit
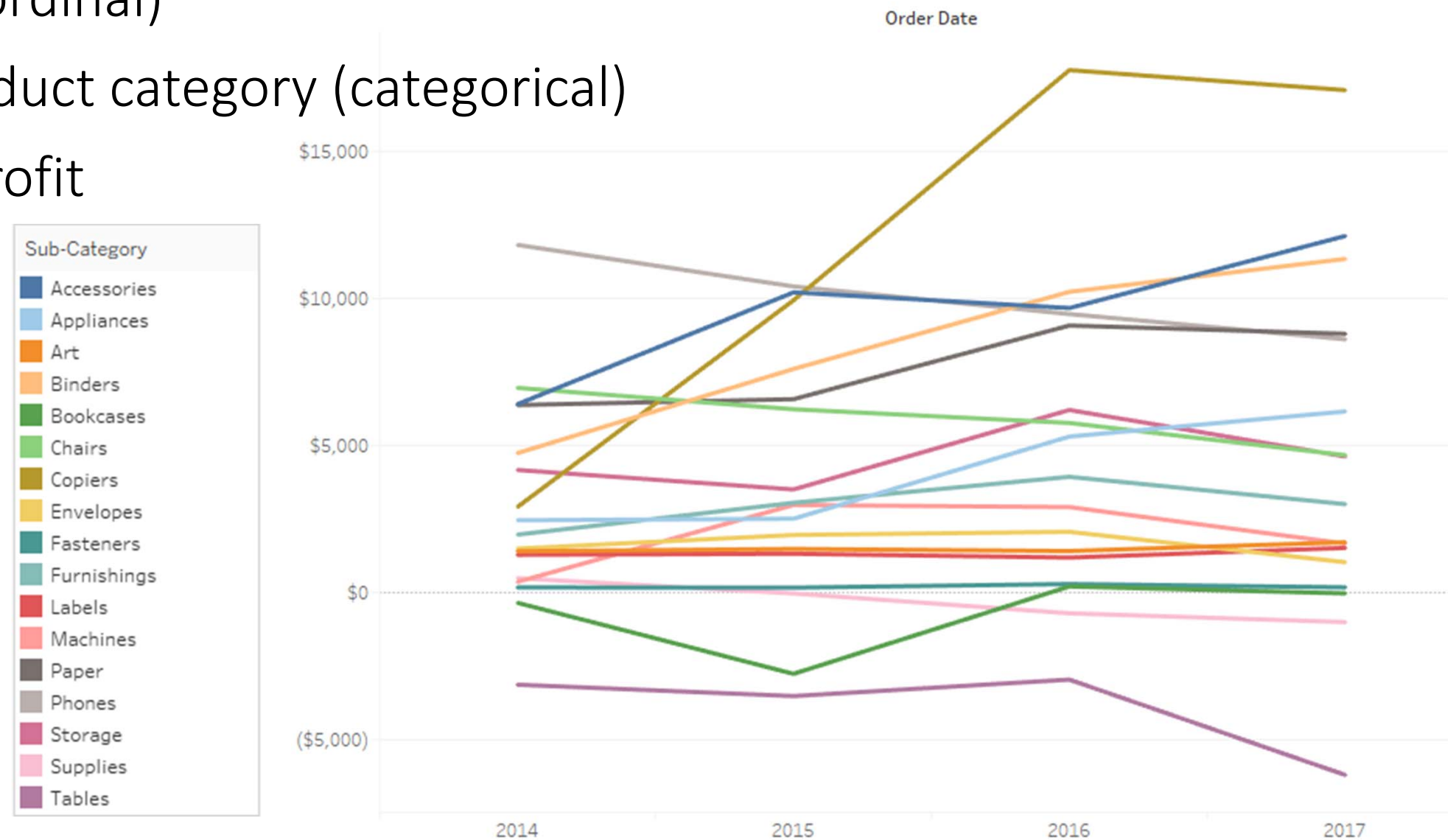
multiple area charts

stacked area graphs

Name Voyager, 2005
http://www.bewitched.com/namevoyager.html

next week

- Monday: Tableau Tutorial I

- Thursday: more about basic visualization techniques
- Why? → Activities and goals that can be supported via visualization
  - What tasks can visualizations support?
  - Munzner, Chapter 3